DOCUMENT RESUME

ED 225 824 SE 040 183

AUTHOR	Dougherty, Knowles; Herbert, Martin
TITLE	Mathematics Applied to Novel Situations (MANS) Test Data. Third Grade. Evaluation Report 4-B-2. Extended Pilot Trial of the Comprehensive School Mathematics Program.
INSTITUTION	CEMREL, Inc., St. Louis, Mo.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Dec 77
NOTE	152p.; For related documents, see SE 040 181-196 and SE 040 340-348.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Educational Research; *Elementary School Mathematics; Evaluation Methods; *Grade 3; Mathematics Achievement; *Mathematics Curriculum; Mathematics Instruction; Primary Education; Program Effectiveness; *Quantitative Tests
IDENTIFIERS	*Comprehensive School Mathematics Program; Mathematics Applied to Novel Situations Test; *Mathematics Education Research

ABSTRACT

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and was funded by the National Institute of Education (NIE). Its major purpose is the development of curriculum materials for the elementary grades. Beginning in 1973, CSMP ran an extended pilot trial of its elementary program. This material looks at whether CSMP students after several years are better able to deal with certain kinds of novel mathematical situations than other pupils. The results of administration of a series of tests called Mathematics Applied to Novel Situations (MANS) are reported. Results indicate classes who have used CSMP curriculum for at least 3 years score better than classes of students who have not. The findings hold for classes at different ability levels, at all testing sites, for both sexes, for right- and left-handers, and for pupils regardless of teachers' judgement of their best learning situation. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED225824

EXTENDED PILOT TRIALS OF THE COMPREHENSIVE SCHOOL MATHEMATICS PROGRAM:

EVALUATION REPORT SERIES

Evaluation Report 4-B-2

Mathematics Applied to Novel Situations (MANS) Test Data

Third Grade

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
CEMREL INC.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 040183

Evaluation Report 4-B-2

Results from the MANS Tests: Third Grade

EPT Year 4

Knowles Dougherty
Martin Herbert

December 1977

Developed by CEMREL, Inc., a private nonprofit corporation supported in part as an educational laboratory by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

Copyright on these materials is claimed only during the period of development, test, and evaluation, unless additional authorization is granted by the National Institute of Education, to claim copyright on the final materials. For information on the status of the copyright claim, contact either the copyright proprietor or the National Institute of Education.

Description of Evaluation Report Series

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials for grades K-6.

Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. The pilot trial is longitudinal in nature; students who began using CSMP materials in kindergarten or first grade in 1973-74, were able to use them in first and second grades respectively in 1974-75, and so on in subsequent years. Hence the adjective "extended".

The evaluation of the program in this extended pilot trial is intended to be reasonably comprehensive and to supply information desired by a wide variety of audiences. For that reason the reports in this series are reasonably non-technical and do not attempt to widely explore some of the related research issues. The list of reports from the first two years of the extended pilot trial is given on the next page. The most comprehensive of these are the following:

1-A-1: Overview, Design and Instrumentation
1-A-3: Final Summary Report, Year 1
2-A-1: Final Summary Report, Year 2
3-B-1: Summary of Second and Third Grade Test Data Year 3
4-A-1: Final Summary Report, Year 4
and 4-A-2: Summary of Grades K-3 Evaluation

Evaluation Report Series

Evaluation Report 1-A-1	Overview, Design and Instrumentation
Evaluation Report 1-A-2	External Review of CSMP Materials
Evaluation Report 1-A-3	Final Summary Report Year 1
Evaluation Report 1-B-1	Mid-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-2	End-of-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-3	End-of-Year Test Data: Standard First Grade Content
Evaluation Report 1-B-4	End-of Year Test Data: CSMP Kindergarten Content
Evaluation Report 1-B-5	Test Data on Some General Cognitive Skills
Evaluation Report 1-B-6	Summary Test Data: Detroit Schools
Evaluation Report 1-C-1	Teacher Training Report
Evaluation Report 1-C-2	Observations of CSMP First Grade Classes
Evaluation Report 1-C-3	Mid-Year Data from Teacher Questionnaires
Evaluation Report 1-C-4	End-of-Year Data from Teacher Questionnaires
Evaluation Report 1-C-5	Interviews with CSMP Kindergarten Teachers
Evaluation Report 1-C-6	Analysis of Teacher Logs
Evaluation Report 2-A-1	Final Summary Report Year 2
Evaluation Report 2-B-1	Second Grade Test Data
Evaluation Report 2-B-2	Readministration of First Grade Test Items
Evaluation Report 2-B-3	Student Interviews
Evaluation Report 2-C-1	Teacher Questionnaire Data
Evaluation Report 2-C-2	Teacher Interviews, Second Grade
Evaluation Report 2-C-3	Teacher Interviews, First Grade
Evaluation Report 3-B-1	Second and Third Grade Test Data Year 3
Evaluation Report 3-C-1	Teacher Questionnaire Data Year 3
Evaluation Report 4-A-1	Final Summary Report Year 4
Evaluation Report 4-A-2	Summary of Grades K-3 Evaluation
Evaluation Report 4-B-1	Standardized Test Data, Third Grade
Evaluation Report 4-B-2	Mathematics Applied to Novel Situations (MANS) Test Data
Evaluation Report 4-B-3	Individually Administered Problems, Third Grade
Evaluation Report 4-C-1	Teacher Questionnaire Data, Third Grade

Key to Indexing

1-C-2 **Observations of CSMP First Grade Classes**

"2" refers simply to the number within a given year and type of data

"C" refers to the type of data being reported
- A: Overview, summary and theoretical reports
- B: Student outcomes
- C: Non-test data

"1" refers to the year of the Pilot Study according to the following:

	Kindergarten	First Grade	Second Grade	Third Grade	. . .
Year 1 (1973-74)	▨	▨			
Year 2 (1974-75)		▨	▨		
Year 3 (1975-76)			▨	▨	

.
.
.

Table of Contents

I. Introduction 1

II. Setting 3

III. Brief Description of MANS/77 5

IV. The Results 9

A. Analysis of Class Means 9

B. Analysis by Student Ability Level 13

C. Other Student Characteristics 15

V. Summary and Discussion 19

Appendices: A. Development of MANS/77 Test Scales 23

B. The MANS Scales 85

C. Statistical Data 101

Introduction

In the fall of 1973, the Comprehensive School Mathematics Program (CSMP) began an Extended Pilot Test (EPT) of its Elementary School Program. Over one hundred teachers began using the program, either in first grade or kindergarten. During the 1974-75 and 1975-76 school years (the second and third years of the EPT) some of these classes continued with CSMP materials in the next grade levels and many new classes began using CSMP materials. The 1976-77 school year constituted the fourth year of the EPT, with some of the original classes continuing with CSMP in the third and fourth grades. Once again, many new classes began using CSMP at grade levels K through 4.

In the evaluation of CSMP student achievement, two major questions have been asked. The first: Do students learn the basic concepts and skills, particularly computational skills, generally expected of students in elementary school? Standardized math achievement tests are used to gather data in answer to this question (Evaluation Report 4-B-1).

The second major question is: Are CSMP students, particularly after two or three years in the program, better able to deal with certain kinds of (novel) mathematical situations than are students who have not studied CSMP? To gather data in answer to this question, a minor study was carried out involving some students in a few classes who were individually tested/interviewed (Evaluation Report 4-B-3) and a major study was carried out involving many classes of students from several CSMP sites who were tested with a specially developed series of tests called Mathematics Applied to Novel Situations (MANS). In this report the results of the administration of these MANS Scales will be presented.

Setting

A series of special tests was developed for administration to third grade classes participating in the CSMP Extended Pilot Test and to designated comparison classes. The process of development for these tests will be described later in this chapter. First, however, the selection of participating classes will be described.

Selection of Classes

The present study was concerned primarily with the cumulative, three-year effects of CSMP, rather than just its effect in the third grade. Therefore, the only third grade CSMP classes eligible were those that had studied CSMP in the first and second grades.

Altogether in 1976-77 there were 81 such third grade classes, located in 17 different sites. The 5 largest sites each containing at least 6 third grade CSMP classes, agreed to cooperate in comparative studies. Two sites later withdrew due to a) illness of a local coordinator and to b) a decision to drop CSMP the following year thereby removing their need for further evaluation data. This left 3 sites:

Site 1 consisted of three suburban school districts of a large Midwestern City. These districts contained schools with various proportions of black and white students and in neighborhoods ranging from lower to upper middle class SES. There were four third grade CSMP classes in each of the three districts, for a total of 12.

Site 2 was a rural/small-town, county school district in the Southeast. The school district was about 60% white, 40% black and was primarily lower middle class. There were 15 third grade CSMP classes.

Site 3 was a relatively small Northeastern City (population approximately 50,000) containing schools with a wide range of SES. There were 6 third grade CSMP classes.

In each of these sites, the local coordinators selected comparison classes from comparable schools in the district, except in Site 2, where it was necessary to use classes from an adjacent school district. As a check on the comparability of the CSMP and comparison classes, scores from an ability or reading test were used in each district as covariates. This allowed adjustments in mean scores to be made on the basis of differences in general ability. Such differences turned out to be small, though the covariates served the very useful purpose of making the eventual comparisons much more precise through a reduction in error variance.

Table 1 provides details of the administration of the covariate test at each site, including mean scores across CSMP and Non-CSMP classes. It can be seen that these mean scores are very nearly equal at each site.

Table 1

Covariate Test by Site

Site Number	Covariate Test	Date of Administration	Administered by	Part of Regular District Testing	Number of Classes		Covariate Mean Raw Scores	
					CSMP	Non-CSMP	CSMP	Non-CSMP
1	Kuhlmann-Anderson	Jan/77	Special Testers	No	12	12	69.7	68.7
2	Comprehensive Test of Basic Skills: Reading Comprehension	May/77	Teachers	No(CSMP Classes) Yes(Non-CSMP)	15	12	27.9	27.8
3	ETS Cooperative Reading Test	May/77	Teachers	Yes	6	12	35.6	36.9

Process of Test Development

The MANS test is the 1977 version of a series of specially developed tests designed to gather data on the major question of this study: Are CSMP students better able to deal with certain kinds of novel mathematical situations than are Non-CSMP students? As in previous years, MANS/77 consists of a series of short scales each one (a) involving rather lengthy directions for administration by the tester to orient the student to the novel task required, (b) consisting of an average of about 9 rather uniform items of increasing difficulty and (c) usually requiring no fixed time limit. The items themselves are generally of the open-ended type, but on a few scales they are the limited choice variety.

The criteria used for creating potential MANS scales are the following. (1) The task required is not covered in most standardized tests. (2) The language and/or content of the task-statement is not relatively more familiar to CSMP or non-CSMP students. (3) The task is recognized as a legitimate application of mathematics. (4) The task is thought to be related to a thinking process contained in the CSMP curriculum.

During the test development process each potential scale was reviewed by the CSMP Evaluation Panel (an external advisory group), the CSMP writing staff and two local school district supervisors who were coordinating CSMP usage in their district.

Through an extensive series of reviews, revisions, pilot trials and data analyses, the scales were revised in order not only to meet the four criteria stated above but also to pass procedural and empirical standards. In order for a scale to be retained, it had to be easily administered by a trained tester and the task easily understood by nearly all students. In additon, the scale had to be free of serious defects: e.g., uniformly easy or difficult items, evidence of much guessing, items in a single scale not testing the same ability, etc.

Appendix A provides a fairly detailed account of this developmental process as it applied to 23 potential scales, 12 of which were finally selected for inclusion in the MANS 77 tests. Two other scales were added to the 12 to make a total of 14: the two being computation scales involving fractions and large whole numbers with which the CSMP students should have been relatively more familiar. The 2 additional scales were included in the administration of the other 12 and the results are reported here as separate scales. However whenever MANS scales in general, or MANS total scores, are referred to, these two scales are excluded. Appendix B gives a reproduction of the actual test booklet pages with a description of the administration for all 14 scales.

The MANS tests (booklets A & B with seven scales each) were administered by specially trained testers, one or two at each site. When two testers were involved in a single site, each one administered as many tests to CSMP classes as to Non-CSMP classes and as a further precaution the testers were not informed as to which of the two types of classes they were testing.

At Site 1, both MANS A and MANS B were administered to all 12 classes.

At Site 2, MANS A was administered to about half the classes in each CSMP school (7 classes) and to about half the classes in each Non-CSMP school (6 classes). MANS B was administered to the other half of the classes, 8 CSMP and 6 Non-CSMP.

At Site 3, MANS A and B were administered to all 6 CSMP classes and to 11 of the 12 Non-CSMP classes (one of these classes was not able to take MANS A).

Thus altogether 69 classes were involved. 25 CSMP and 29 Non-CSMP classes took MANS A: 26 CSMP and 30 Non-CSMP classes took MANS B.

Brief Description of MANS/77

Although a more complete presentation of the MANS scales is presented in Appendix B, in this section there is a very brief description of the tests and each of the scales. There were 14 scales in all: Scales A1-A7 in Booklet A and Scales B1-B7 in Booklet B. Each of Booklets A and B were 16 newsprint pages and each required a total of 50-60 minutes administration time by a specially trained tester.

Within each booklet the format was the same. For each scale the tester explained what the required task was and did one or two samples with the whole class, all according to a carefully designed script. Then the students went ahead and did the test items on that scale, which involved one or more pages. They did not turn the page to begin a new scale until they were told to do so by the tester - then everyone turned at once - and the same procedure was followed by the next scale.

Thus for any given scale, each student in the class had the same amount of time. For most scales this time limit was fairly flexible, with a suggested time limit for the tester, which could be extended if students were still working. For some scales which involved estimation, and in which laborious calculations were discouraged, strict time limits were adhered to.

For each scale the following information is given: (1) the scale letter/number, its name, and the number of items, (2) an abstract of the content, and (3) a relatively easy sample item. There is an asterisk placed before the two scales not of the MANS-type.

SCALE A1: Height and Weight Table (6 items)

ABSTRACT: Read and interpret data from a table of students' weights and heights for two different years

SAMPLE: Who stayed the same height? (Note to the reader: the table can be seen in Appendix B)

SCALE A2: Estimation (25 items)

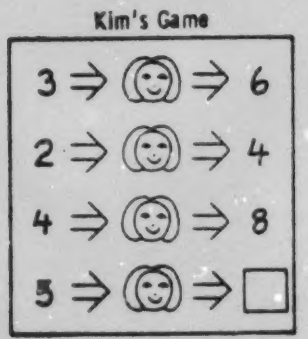
ABSTRACT: Quickly estimate which of 5 standard intervals contains the answer to each of a series of computation problems. Three separate pages containing 8 addition, 8 subtraction and 7 multiplication problems respectively.

SAMPLE: 100 - 93 0 10 50 100 500 1000

SCALE A3: Functions (8 items)

ABSTRACT: For each of several problems, determine from 3 pairs of numbers what the "secret rule" is which produces the second number from the first, and use it to find the missing number from the 4th pair.

SAMPLE:



SCALE A4: Two Stage Word Problems (5 items)

ABSTRACT: Word problems (printed in booklet and read by tester) in which two different operations must be performed and where the numbers in the given data are relatively small.

SAMPLE: Our hens lay 9 eggs every day.

Each day we eat 6 of them and give the others away.

During the next 5 days how many eggs will we give away? _____

*SCALE A5: Large Number Computations (12 items)

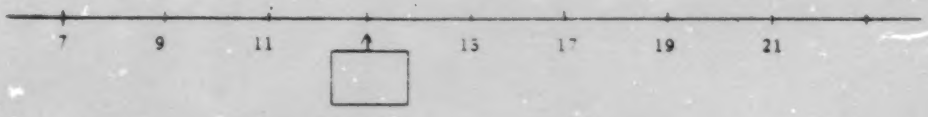
ABSTRACT: Solve computation problems given in an open sentence format, with the boxes sometimes in non-standard positions, and with numbers in the hundreds but relatively easy to work with (addition, subtraction and multiplication).

SAMPLE: 500 + ☐ = 800

SCALE A6: Number Line Labelling (8 items)

ABSTRACT: Label the indicated "mark" on several number lines, where marked intervals vary from item to item and where other marks are irregularly labelled.

SAMPLE:



SCALE A7: Hints and Problems (5 items)

ABSTRACT: Quickly complete a given addition problem by using the answer to another problem where one addend is the same as, and one is only slightly different from the given problem.

SAMPLE: Hint: 537 + 293 = 830

537 + 283 = ☐

SCALE B1: 2 or 5 or 10 (10 items)

ABSTRACT: Quickly estimate whether a given number is about 2 or 5 or 10 times as large as another given number.

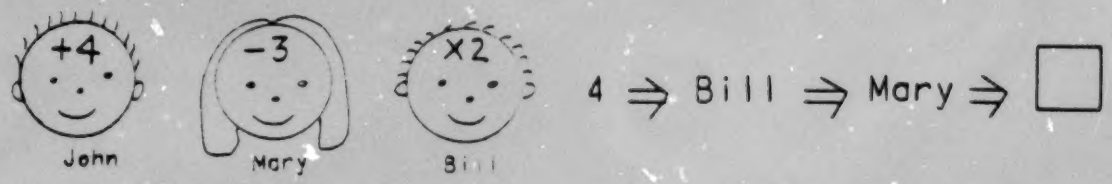
SAMPLE: 60 is about _____ times as large as 31

12

SCALE B2: Composite Functions (9 items)

ABSTRACT: Starting with a given number, apply one or more operations in sequence and determine final result. Also, same process except final result is known and starting number is to be determined.

SAMPLE:



SCALE B3: Equation Fluency

ABSTRACT: Given the symbols: = + - x 1 2 3 ( ), construct as many different equations as possible

SAMPLE ANSWER: 3 - 1 = 2

SCALE B4: Circle the Larger (13 items)

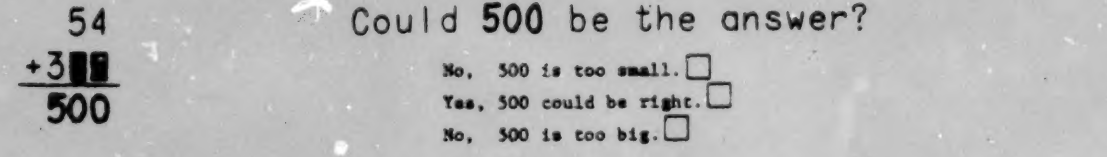
ABSTRACT: Given pairs of computation problems, quickly determine which one has the larger answer.

SAMPLE: 371 + 248 370 + 258

SCALE B5: Missing Digits (8 items)

ABSTRACT: Given a computation problem with one or two digits of the problem crossed out, determine whether or not the given answer could have been right (before the digits were crossed out).

SAMPLE:



SCALE B6: Word Problems with "Rounding" (5 times)

ABSTRACT: Solve word problems (printed in booklet and read by the tester) involving division in which the given numbers do not divide evenly - i.e., the answer, which must be an integer, can be obtained by rounding the obtained quotient up or down. The numbers of the given data were relatively small.

SAMPLE:

An elevator can't hold more than 5 people.

23 people want to ride to the top floor.

How many times will the elevator have to go up? _____

SCALE B7: Fractions (8 items)

ABSTRACT: Solve problems of the form x of y = ☐ or x of ☐ = y where x is 1/2 or 1/3.

SAMPLE: $\frac{1}{2}$ of 12 = ☐

Results

In this section the third grade MANS/77 results will be presented from several different points of view. Since the curriculum (CSMP or otherwise) is imparted via an individual teacher to a class as a whole, the most valid unit of analysis for comparison is the class. By using classes rather than students as the unit of analysis, the number of units is greatly reduced, with a resultant increase in the differences required to show statistical significance. The analysis of class means is presented first, as the most powerful and conservative argument for differences between CSMP and Non-CSMP.

Analysis of Class Means.

The results are summarized in Table 2 on the next page. The basic data are the adjusted class means - adjusted on the basis of the covariate score. Because the CSMP and Non-CSMP classes were always very close in ability, as measured by the covariate score, these adjustments at any site were never greater than 0.1 for any subtest.

The right hand column of Table 2, the combined results, shows that CSMP classes scored significantly higher than Non-CSMP classes on both the "total" MANS/A and the "total" MANS/B. Looking at the individual scales (excluding A5 and B7), it can be seen that CSMP classes scored higher than Non-CSMP classes in 11 of them, significantly so in 8 of the 11. As expected, there was also a significant difference in favor of CSMP on both Scales A5 and A7.

From the same Table 2, these across-site results can be checked on an individual site-by-site basis. On the "total" MANS/A and "total" MANS/B, the across-sites results are corroborated in Site 1 and Site 3. In Site 2, the results are still in favor of CSMP but not significantly so.

Table 2

Adjusted Mean Scores, MANS Scales

First Entry: CSMP Classes, Second Entry: Non-CSMP Classes

*, **: Significant at .05, .01 levels respectively [4]

MANS A	Site							
	One		Two		Three		Combined [3]	
Number of Classes	12	12	7	6	6	12	25	30
Raw Score, Covariate [1]	69.5	68.6	28.0	28.5	35.6	36.7	54.5	55.9
A1. Height and Weight Table	3.5	3.4	2.9	2.9	3.3	3.7	3.3	3.4
A2. Estimation	10.7	10.3	8.7	8.7	11.7	9.9	10.5	9.7 *
A3. Functions	4.4	3.4 **	2.9	2.5	4.9	3.8 **	4.2	3.4 **
A4. Two Stage Word Problems	3.1	2.6 **	2.7	2.2	3.2	3.1	3.0	2.7 **
A5. Large Number Computations	6.1	5.3 **	4.6	4.4	6.4	5.3 **	5.9	5.0 **
A6. Number Line Labelling	4.8	4.6	4.0	4.0	5.4	4.9 *	4.8	4.5
A7. Hints and Problems	2.1	1.8	1.3	1.3	2.2	1.7 *	2.0	1.6 **
Total minus A5 [2]	28.5	26.2 **	22.5	21.7	30.9	27.1 **	27.8	25.3 **
MANS B								
Number of Classes	12	12	8	6	6	11	26	29
Raw Score, Covariate [1]	69.5	68.6	28.0	28.0	35.6	36.7	54.5	56.7
B1. 2 or 5 or 10	6.3	5.8 *	5.3	5.2	6.4	6.1	6.1	5.7 *
B2. Composite Functions	6.4	5.5 **	5.0	4.4	7.3	5.7 **	6.3	5.3 **
B3. Equation Fluency	6.0	6.1	5.2	4.8	6.5	6.2	5.9	5.8
B4. Circle the Larger	9.3	7.9 **	8.0	7.5	9.3	8.2 **	8.9	7.9 **
B5. Missing Digits	3.7	3.0 **	2.9	2.9	3.0	3.1	3.3	3.0 **
B6. Word Problems "with Rounding"	2.3	2.2	1.9	1.6	2.5	2.4	2.3	2.1
B7. Fractions	4.1	2.8 **	2.9	2.0 **	4.8	3.3 **	4.0	2.8 **
Total minus B7	34.0	30.5 **	28.2	26.4	35.0	31.7 **	32.8	29.8 **

[1] Kuhlmann Anderson, CTBS Reading Comprehension and ETS Reading for Sites 1,2,3 respectively. Transformed percentile ranks for Combined.

[2] Since tests A5 and B7 were somewhat different from the other MANS Scales, in that they were straight computation, the total scores on A and B were computed without these two.

[3] Because a different covariate score was used at each site, the raw scores were transformed into percentile ranks that were comparable from site to site; i.e, site-to-site differences were reduced as much as possible in order to concentrate on CSMP/Non-CSMP effects.

[4] Significance tests were based on a one-way Analysis of Covariance procedure, with 1 and 21, 1 and 10, 1 and 15, and 1 and 52, degrees of freedom for Sites 1, 2, 3 and combined respectively.

In order to look more closely at the MANS results at the scale level, a somewhat more sophisticated approach was used. For each scale at each site two statistics were calculated: (a) the difference in the adjusted means between CSMP and Non-CSMP classes and (b) the pooled standard deviation of class scores after adjustment for covariate (the square of this statistic is simply the reduced error variance in the analysis of covariance). The first statistic divided by the second gives the mean difference in standard deviation units. This allows a more realistic comparison of differences across sites.

For purposes of classification, when this ratio was equal to or greater than 0.9, this was called a "large effect", i.e., nearly one standard deviation or more. Across the three sites, 20 of the 42 differences represented large effects. The other 22 differences were equal to or less than 0.6 and might be called moderate or small effects.

Table 3 shows the scales presented in four groups according to whether there are large effects: (a) in all three sites, (b) in only two sites, (c) in only one site, and (d) in none of the sites. All the large effects are in favor of CSMP classes.

Table 3

Scale by Scale Analysis of Group Differences

Number of Sites Showing Large Effects	$\frac{X_1-X_2^*}{Sd_r}$				MANS Scales
	Site 1	Site 2	Site 3	All Sites	
All Three	(1.2)	(2.1)	(2.6)	1.9	Fractions (B7)
	(0.9)	(0.9)	(1.8)	1.4	Circle the Larger (B4)
	(1.5)	(0.9)	(2.1)	1.3	Functions (A3)
	(1.2)	(0.9)	(2.0)	1.1	Composite Functions (B2)
Two of the Three	(1.3)	(0.9)	0.3	0.8	Two Stage Word Problems (A4)
	(1.8)	0.4	(2.0)	1.4	Large Number Computation (A5)
One of the Three	0.6	0.1	(1.4)	0.8	Hints and Problems (A7)
	0.4	0.0	(1.5)	0.7	Estimation (A2)
	(1.0)	0.0	-0.3	0.6	Missing Digits (B5)
	0.4	-0.2	(1.4)	0.5	Number Line Labeling (A6)
None of the Three	0.6	0.1	0.5	0.6	2 or 5 or 10 (B1)
	0.2	0.6	0.1	0.2	Word Problems & Rounding (B6)
	0.1	0.4	0.3	0.1	Equation Fluency (B3)
	0.2	0.2	-0.7	-0.2	Height and Weight (A1)

*X_1 = Adjusted mean score across CSMP classes

X_2 = " " " " Non-CSMP classes

Sd_r = pooled standard deviation of adjusted score (Sd_r^2 = reduced error variance in analysis of covariance)

In comparing Tables 2 and 3, one notes that, for Site 2, 6 scales show a large difference (Table 3) while only one of them shows a statistically significant difference (Table 2). This is due partly to the fact that for any one scale there were fewer classes at Site 2 and partly due to the fact that the covariate was much less effective there in reducing error variance.

It is interesting to compare these results to those on similar kinds of tasks used in the testing in previous years of the EPT. (1) In previous years, items dealing with fractions and large numbers were contained in several different scales with the item analysis indicating that CSMP students had a decided advantage. This year, with the two isolated in separate scales (B7 and A5), we see that CSMP students did significantly better. (2) It was not surprising to see that the two scales involving functions (A3 and B2) showed large effects in all sites. Previous scales involving relational thinking have usually produced results in favor of CSMP. However, differences this year are greater than in previous years (grades one and two). (3) Two Stage Word Problems (A4) showed large effects in two of the three sites, while Word Problems with Rounding (B6) did not. This is typical of previous years. There is no consistent finding for word problems in general; the results depend on the specific kinds of word problems involved.

In general there were larger differences in favor of CSMP on the MANS this last year than in previous years. The evidence over three years indicates that as classes study more CSMP, their advantage over Non-CSMP classes increases on tests like the MANS.*

In Appendix C, considerable data is presented on each of the MANS scales. Item analysis data is presented along with item and scale correlations, reliability measures and scores by ability level. Also given for each scale is a graph plotting each third grade class according to mean score for the scale and the percentile rank on the covariate. By examining these graphs, one can check to see whether or not the above results are largely due to differences in CSMP and Non-CSMP classes at just one general ability level. It will be seen that, although there is a narrow middle range of ability levels which is not well represented by either CSMP or Non-CSMP classes, the results are reasonably consistent across all ability levels of classes participating in the study.

*In a related study, some additional evidence for this latter generalization has been gathered from an elementary school serving an upper middle class suburb of a small city in upstate New York. Two classes in this school had studied CSMP through second grade (1973-76); the other two had not. The following year (1976-77), these students were thoroughly mixed into four third grade classes. For each class therefore, half the students had studied CSMP through second grade and the other half had not. In the Spring of 1977 MANS/A was administered to two of the four classes, and MANS/B to the other two with the following results:

		Adjusted Means Total MANS/A	Adjusted Means Total MANS/B
Classes 1 & 2:	Veteran CSMP Students	43.9	
	New-to-CSMP Students	42.8	
Classes 3 & 4:	Veteran CSMP Students		52.1
	New-to-CSMP Students		50.1
Differences in the Adjusted Means		1.1	2.0

The differences of 1.1 and 2.0 can be compared to differences of 3.0 and 4.5 for CSMP versus Non-CSMP students of comparable ability in Sites 1, 2 and 3. Thus we see that the more years of CSMP experience students have had, the higher they score on MANS tests.

Analysis by Student Ability Level

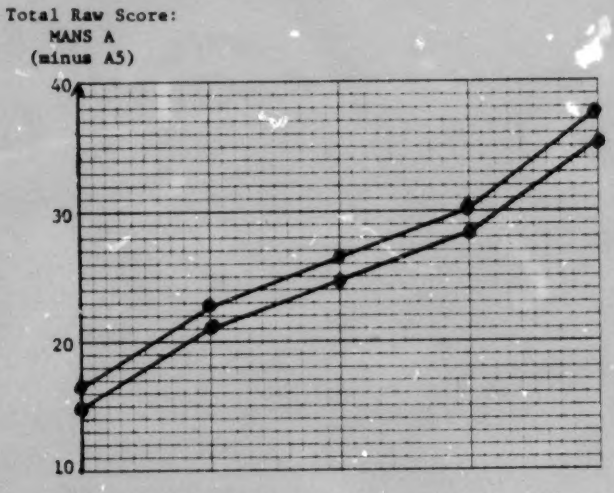
One can legitimately ask if the result obtained using class means as the unit of analysis will hold when data is aggregated across individual students. For example, is it also true that if one looks at low ability students only, roughly the same group differences will appear as with high ability students? To answer this question the following procedure was used.

Based on the covariate score, each student was assigned to either the first quintile (for students below the 20th percentile rank), or the second (20th-40th) or third (40th-60th), or fourth (60th-80th) or fifth quintile (above 80th percentile rank). Then the mean score for all CSMP students in each quintile was computed and the same procedure used with Non-CSMP students. These mean scores are shown in numerical and graphical form in Figures 1 and 2, next page.

As can readily be seen, the CSMP advantage is consistent across each of the five ability levels for both the MANS tests. In Appendix C one can check the above information for each individual scale. Upon examination, it will be found that the consistent pattern, illustrated above, holds true for each of the eight scales in which there is a significant difference. In addition, in the three scales where the CSMP score is higher but not significantly so, in no quintile do Non-CSMP students outscore their CSMP counter-parts. Clearly it is the case that the general results hold true for all ability levels.

Figure 1

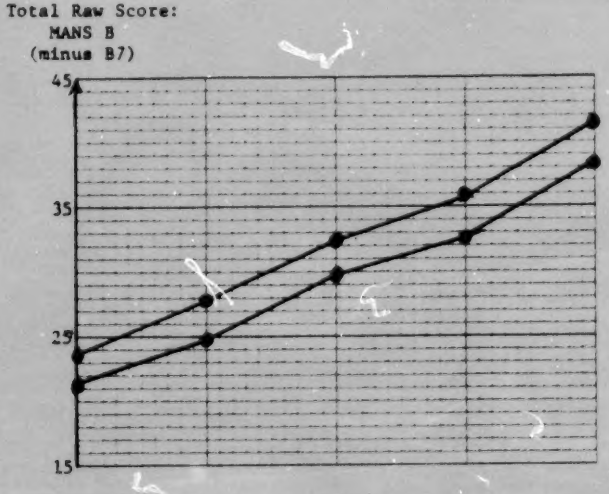
MANS/A Performance by Covariate Quintiles



Range of Percentile Ranks:	under 20	20-40	40-60	60-80	over 80
Mean, CSMP Students:	16.4	[illegible]	26.5	30.2	37.8
Mean, Non-CSMP Students:	14.9	21.0	24.6	28.2	35.4
Pooled Standard Deviation:	6.4	7.9	8.8	8.9	8.2

Figure 2

MANS/B Performance by Covaraite Quintiles



Range of Percentile Ranks:	under 20	20-40	40-60	60-80	over 80
Mean, CSMP Students:	23.7	27.9	32.5	35.8	41.3
Mean, Non-CSMP Students:	21.1	24.8	29.9	32.7	38.3
Pooled Standard Deviation:	7.2	7.6	7.9	8.1	7.0

Other Student Characteristics

While classes were being given the MANS tests, the classroom teachers were asked to complete a form in which they were to classify each student along each of four dimensions: (a) sex (male or female), (b) handedness (right or left), (c) mathematical problem solving ability (high, above average, average, below average and low), (d) best learning situation (learns best in a group instruction format, learns best when working by himself (herself), both situations about the same).

Table 4a, b, c, and d show the MANS results for each student category within the four dimensions. For the sake of simplicity, the scores on all 12 MANS-type scales are compiled for a total MANS score. Since slightly different groups took MANS/A and MANS/B the figures given for "number of students" and "covariate score" are the averages across the two groups.

From the four-part Table 4 it can be seen that CSMP students scored higher than Non-CSMP students in every single student category. It is noteworthy that in all categories the numbers of students are quite high. Apart from the finding that CSMP students scored higher than Non-CSMP students in every classification, it is worthwhile noting some other findings which are rather interesting even though they do not pertain to CSMP/Non-CSMP comparisons:

(1) On the sex dimension (Table 4a), regardless of curriculum, boys outscore girls on the MANS. This result is all the more interesting when it is noted that girls outperformed boys on the covariate and on the tests of standard math achievement.* It is worth noting that each of these generalized sex differences are consistent across sites.

(2) On the matter of handedness (Table 4b), it is noteworthy that left-handers outscored right-handers regardless of curriculum. Since the covariate scores for the two groups are very similar, these results cannot be easily attributed to differences in general ability.

(3) With the teacher-judged dimension of math problem solving ability (Table 4c), the most interesting additional finding is the differences in scores between the two highest group of students. As one would expect, the highest teacher-rated group out performs the above average group, for CSMP. But for Non-CSMP, this is not the case. By checking the covariate scores, we see that the phenomenon is repeated, but not as dramatically: i.e., "High" Non-CSMP students do have a higher covariate score than "Above Average" Non-CSMP students. In order to verify this phenomenon, the scores on the standarized tests of traditional math achievement were compared for these teacher-judged groups. In the CTBS Computation as well as Concepts and Applications (Site 1 and 2) and in the ETS Cooperative Primary Math (Site 3) the phenomenon is repeated almost exactly. It appears that CSMP teachers are able to distinguish between high and above average math problem solvers and Non-CSMP teachers are not. Perhaps it is the case that the CSMP curriculum provides the opportunity for teachers to witness the distinction between the math behavior of superior students versus those who are only above average.

*On the standard math achievement tests, combining CSMP and Non-CSMP students, the means were as follows:

	CTBS Computation	CTBS Concepts and Applications	ETS Cooperative Primary Mathematics Test
Boys	34.6	31.6	41.7
Girls	37.3	33.9	42.0

Table 4a
MANS Results by Sex

	Number of Students		Covariate Score		Total MANS Score	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Boys:	256	298	66.0	66.2	60.1	56.1
Girls:	259	292	68.9	71.1	57.8	54.5

Table 4b
MANS Results by Handedness

	Number of Students		Covariate Score		Total MANS Score	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Right Handed:	487	544	66.7	67.3	58.8	54.4
Left Handed:	54	74	67.2	66.9	60.1	56.8

Table 4c
MANS Results by Math Problems Solving Ability
(as judged by teachers)

	Number of Students		Covariate Score		Total MANS Score	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
High:	55 (10%)	74 (12%)	81.9	79.3	83.2	68.9
Above Average:	105 (19%)	135 (22%)	75.6	76.7	73.8	68.3
Average:	210 (38%)	234 (38%)	69.3	67.6	58.9	53.2
Below Average:	99 (18%)	105 (17%)	57.5	57.0	46.1	43.3
Low:	83 (15%)	74 (12%)	49.7	52.1	37.5	35.7

Table 4d
MANS Results by Best Learning Situation

	Number of Students		Covariate Score		Total MANS Score	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
In group Instruction Format:	108 (20%)	148 (24%)	64.6	63.3	55.9	50.2
When Working by Himself (Herself):	124 (23%)	99 (16%)	59.9	58.1	64.6	59.7
Both Situations about the same:	307 (57%)	365 (59%)	68.1	69.0	49.4	44.9

(4) In reference to the "best learning situation" dimension (Table 4d), it is not at all clear why the scores (covariate as well as MANS) ranked in the following order: both situations about the same, learns best in group instruction format, learns best when working by himself (herself). On the surface, it would seem that a student who learns best by working alone might well be more able than other students. Apparently, either that notion is false or the teachers had a different understanding of the categories. Perhaps the teacher generally interpreted the categories as: "can learn in any situation", "needs work with the teacher in a small group", "needs individual attention".

Summary and Discussion

On the tests of Mathematics Applied to Novel Situations there is considerable evidence to indicate that classes of students who have used the CSMP curriculum for at least three years score better than classes of students who have not. The finding holds for the classes in general, for classes at different ability levels, and for classes in all three sites tested (but only significantly so in two of the three sites). It is also clear that this finding holds at each ability level of the students tested, whether ability is measured by objective, standardized tests or by teachers' judgment of math problem solving ability. In addition the finding holds for boys and girls, for righthanders and lefthanders, and for students regardless of teachers' judgment of their best learning situation.

At this point several issues will be explored.

1. Fairness of the MANS Scales

The MANS scales are produced by people closely associated with the CSMP program and are intended to test for application to new situations of skills supposedly covered in the CSMP curriculum. This is virtually a no-lose situation. It may be argued that it is possible to develop scales on which the CSMP students are less able to apply their knowledge. This, of course, cannot be absolutely disproved. It is true however that standardized test data which if anything should favor Non-CSMP classes has in general revealed no over-all advantage one way or the other. The correlation of the MANS total scores with standardized math scores is fairly high, in the 60's and 70's, making the performance of CSMP students on the MANS test even more impressive. And it is true that in curriculum evaluation, significant differences in these areas (applications and transfer tasks) are relatively hard to obtain.

2. Alternative Explanations

Given that the results reflect real differences in performance between CSMP and Non-CSMP classes, it is necessary to ask whether these differences can be attributed entirely to the CSMP curriculum, or whether other explanations are as salient. This is always a problem when randomization of students to CSMP and Non-CSMP is not possible as was the case in these studies where CSMP third grade classes already existed and where comparison classes were located and tested after the fact.

The assignment of students to CSMP or to Non-CSMP classes, probably was not a factor as the following analysis shows. There were a total of 13 schools with participating third grade CSMP classes and in 11 of these schools it was the case that every third grader in the school was studying CSMP. In 12 of the 15 Non-CSMP schools a similar statement could be made. In effect then, for the most part, assignment to curriculum (CSMP or Non-CSMP) was effective at the school level, rather than at the student or even class level.

The above paragraph shows that the alternative explanation of superior CSMP teachers causing better performance, a possibility raised frequently in past Evaluation Reports, is much less tenable now. When some classes in a school got CSMP and the others were comparison classes, one had to question whether or not the better teachers, by assignment or volunteering, usually got the CSMP classes. But since the school is the unit of selection this alternative explanation requires that the teachers in the CSMP schools were systematically better than those in the Non-CSMP schools, an unlikely supposition in view of the apparent similarities in school SES levels and student ability levels. (It may have been true that when CSMP was begun in a school, at the first grade level, the first grade teachers at that school were judged by the local coordinator to be better able to teach CSMP than the first grade teachers in another shcool. Even if this were the reason for one school using CSMP and another not, the effects of different first grade teachers would have been largely dissipated by the end of third grade.)

As a further check on CSMP versus Non-CSMP teacher comparability, all participating teachers in Site 1 were interviewed regarding their attitude towards teaching math and the instructional time allowed for mathematics. The results are shown in Tables 5 and 6.

Table 5

Attitude Toward Teaching Math

Teacher Attitude	Number of CSMP Teachers	Number of Non-CSMP Teachers
Math one of her favorites	4	4
Likes or enjoys teaching math	5	5
Average, so-so, no answer, etc.	3	3

Table 6

Time Spent teaching Math

Time Spent	Number of CSMP Teachers	Number of Non-CSMP Teachers
Spends less than 40 min/day	3	2
Between 40 and 55	4	6
Spends an hour or more	5	4

As one can readily see, there is little difference between CSMP and Non-CSMP teachers on these two dimensions.

The CSMP program is being taught for the first or second time by the CSMP teachers in this study and hence there may exist a generalized Hawthorne effect; teachers try harder with anything new or different. But this seems to be a rather weak argument, especially in view of previous data (Evaluation Report 2-B-2) which showed no such effects.

None of these alternative explanations stand up well to careful scrutiny and one concludes that it is the CSMP curriculum that is making the difference.

3. Generalizability of Results

Given that the results reflect real differences in favor of CSMP classes, and that the CSMP curriculum made the difference, to what extent can one generalize these results to school districts other than the 5 which participated in this study? These school districts were not, as one would have liked, stratified samples of the target population of school districts according to size, SES, location, etc. They were merely 5 of the largest users of the third grade CSMP materials and they agreed to cooperate in collecting comparative data.

It is fortuitous that between them they are broadly representative of many school districts. Although one wished that the study could have included a large urban school district with inner city classes, for example, two previous comparative studies from earlier grades did involve such school districts with results very favorable to CSMP. Thus while one can never properly address this issue statistically without a true sampling procedure, it is entirely reasonable to review these results in comparison to previous studies, with other groups of students at lower grade levels. An effort such as this would cover a lot of ground and is done in Evaluation Report 4-A-2. We summarize here by saying that these results are consistent with previous data.

Appendix A

Development of MANS 77 Test Scales

Historical Antecedents

MANS/77 has been developed as part of the Extended Pilot Trial of CSMP materials. This evaluation effort has included a rather extensive amount of testing of students within the CSMP curriculum as well outside it. The testing program was outlined from the beginning (p.33, Evaluation Report 1-A-1) in the form of three questions:

> Do students learn the basic concepts and skills, paricularly computational skills, generally expected of students in elementary school?
>
> Do students learn the specific skills and concepts of the CSMP curriculum or, otherwise stated, do they attain behavioral objectives of the program?
>
> Are CSMP students, particularly after two or three years in the program, better able to deal with certain kinds of (novel) mathematical situations than are students who have not studied CSMP?

Special tests have been created each year of the EPT to help answer the third question:

In EPT Year One, they were called "First Grade Tests of General Cognitive Skills". These tests were designed to measure students' skills in classification, relational thinking and analytic reasoning and were administered individually to each student. The results are given in Evaluation Report 1-B-5.

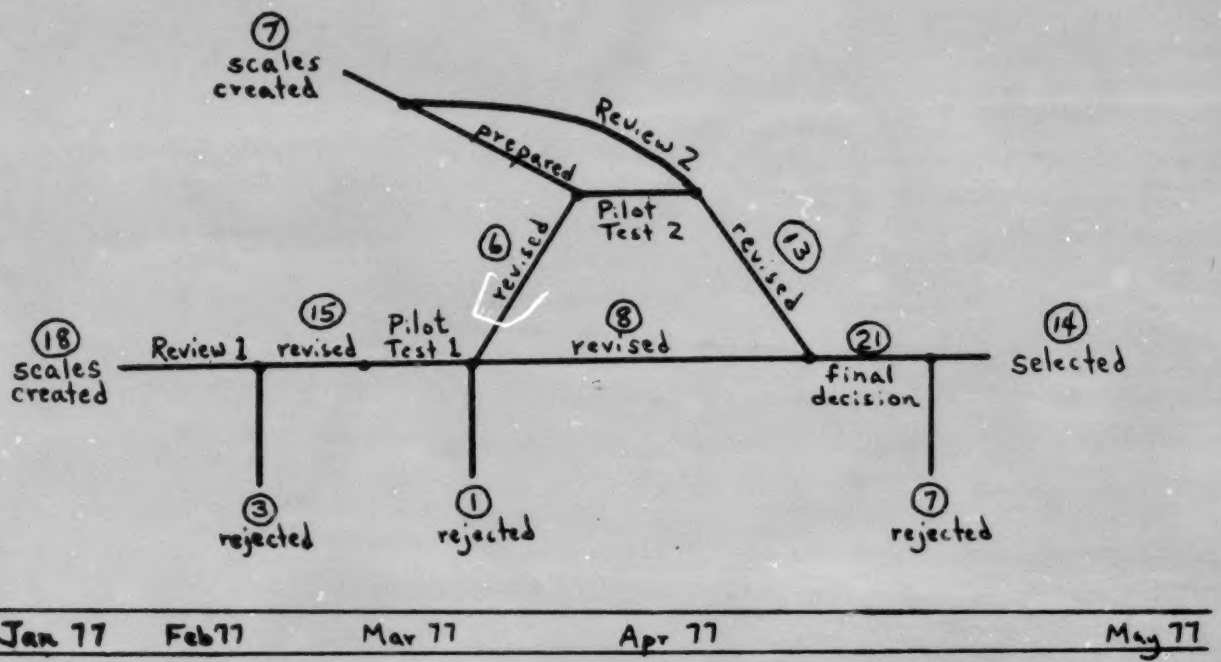
In EPT Year Two, they were called "Second Grade Comparison Tests". There were 14 subtests, 9 group administered and 5 administered individually. (See Evaluation Report 2-B-1).

In EPT Year Three (1975-76), the measure was called "Second Grade Mathematics Applied to Novel Situation Tests", (MANS). There were ten subtests, all group administered (Evaluation Report 3-B-1).

In EPT Year Four (1976-77), MANS/77 was developed as the grade three equivalent of MANS/76.

Process of Development of Scales

The process of development is shown schematically below. Each circled entry represents the number of scales involved.



Creation of Scales

It can be seen that (18) scales were produced originally and (7) more were produced later. Some of these were revisions of previously used scales. Others were newly created, either directly by the Evaluation staff or based on suggestions by the Evaluation Panel (which met in January) or by the CSMP writers (who met with the evaluators in January and February and who were asked to submit sample items). The criteria used in creating potential MANS scales were the following:

(1) The task required of the student is not already covered in standardized tests.
(2) The language and/or content of the task-statement is not relatively more familiar to CSMP or non-CSMP students.
(3) The task is one that is recognized as a legitimate application of mathematics.
(4) The task is thought to be related to the thinking process contained in the curriculum.

Review 1

On 25 February 1977, the 18 scales were sent for comment to the Evaluation Panel, to two local CSMP coordinators (employees of St. Louis area public school systems), and to the CSMP writers.

By March 7th comments were received from the 13 reviewers. Based on these reactions, three of the scales were dropped and the remaining 15 scales and directions to students were revised. Three sets of five scales each were then prepared for pilot testing.

Pilot Test 1

Between March 11th and 16th, these scales were tried out with third grade classes (six in all) in two lower middle class suburban elementary schools, one predominantly white, the other predominantly black. Each set of 5 scales was pilot tested in three classes altogether. In a few cases, where time permitted, revisions were made between administrations.

During the period March 16 to 24, the results were considered, taking into account both (a) the reactions of testers and observers (including the classroom teachers) to the test administrations and (b) the subsequent analysis of the data. This analysis included such things as indices of test reliability, item difficulty, item discrimination, distribution of scores, items not reached or omitted, and correlation with teacher ratings. Based on criteria outlined on the next page of this report, one of the 15 scales was dropped altogether. The other 14 were revised: six were scheduled for a further pilot test, and eight were deemed available for use.

Review 2

During this same period, seven additional scales were created, five of which were distinctly different types of word problems. These were then sent to the reviewers.

Pilot Test 2

Because of time pressure, it was necessary to have reviewers comment on the 7 new scales at the same time as these 7 plus 6 revised scales from the first round were being pilot tested (March 25 to 31). This second round of pilot tests was conducted with a new group of six third grade classes. Three were from another school in one of the above districts and the other three were from a small rural district, and unlike the other pilot classes, were using the CSMP program. The thirteen scales were generally tested with two or three classes each.

From April 1st to 7th the comments from the reviewers were received on the additional seven scales and the results of the second round of pilot testing were analysed including the observations of the testers and observers. All 13 scales piloted in this second round were revised and deemed available for use.

Final Selection

Thus altogether there were 21 revised scales available for use; 8 from the first round of pilot tests and 13 from the second round. Of these 21 scales, 14 were selected by the Evaluation staff for inclusion in the final version of MANS 77. The criteria used for rejecting a scale at this and previous stages are described below.

(1) Analytic. If several reviewers have reservations about the scale or if one or two have serious reservations, the potential scale will be dropped.

(2) Procedural. If the pilot tests show a scale to be too difficult to administer, it will either be greatly revised or dropped. Procedural difficulties include such things as (a) the instructions take too much time (b) the demonstration materials (if any) are too cumbersome, (c) too many students apparently do not understand what the task is, etc.

(3) Empirical. If serious defects become apparent in the analysis of the data from the pilot trials, a scale will either be greatly revised or dropped. Serious defects include such things as (a) many students got all or almost all the items right (or wrong), (b) the items appear not to be testing the same ability (low reliability), (c) the items are not in order of increasing difficulty, or there is a sudden leap in the level of difficulty, (e) many items were omitted by large numbers of students, (f) many students did not complete the test, (g) there was evidence of guessing, etc.

The above factors were taken into account when the final selection was made of 14 scales from the surviving 21. But other considerations were also used. For example, if another potential MANS scale appeared to be testing the same ability, and in other respects appeared to be superior, the less superior scale may be excluded from the final version of the MANS in order to create a balanced test. It is also quite possible that a good potential MANS scale fails to make the final version because of its time length, or because there simply were so many other good scales that were better.

By the end of the school year, the scales had been printed (two booklets of 7 scales each), 4 testers trained and the tests administered to about 70 classes.

The 25 Potential MANS Scales and "How They Lived or Died"

The remainder of this present discourse will be devoted to a consideration of each potential MANS scale in turn. "Initial Version" will mean a potential MANS scale ready for comment by the reviewers. If the scale is to be timed, the time allotment is given in the final verison. If no allotment is given, it means that the scale is given a flexible time limit, a time limit after which the tester uses her/his discretion about terminating the students' work.

Before going into the individual histories of each of the 25 scales, a brief discussion of word problems in general is given. The word-problem-type scales actually originated from a more general inquiry about the inclusion of scales involving word problems. This query was sent to reviewers as part of the first round of scales.

Word Problems

From the start it was not clear what types of word problems, if any, should be included in the MANS/77. The reviewers were asked to address themselves to question, at the same time reminding them that (a) word problems were included in last year's MANS/76 (2nd graders) and (b) fairly straightforward types of word problems already occur in the standardized tests that students take.

MANS/76 Word Problems. There were 8 series of pictures on two consecutive pages of the student booklet; for each series of 3 pictures there was a story at the end of which was a question the student was expected to answer. The questions were written directly below each series of pictures. The tester read each story and question aloud two times. Below is an example.

Tester's Script

"First picture: 'Four children each get the same allowance from their mother.'
Second picture: 'The four children put their allowances together.'
Third picture: 'They have altogether 12 dollars'
'How much did each child get?"

Student's Booklet

Reactions of the Writers, Panelists and Coordinators. The concensus of all reviewers was that 'word problems should be included if only they could ..." The following are the substantive qualifiers each brought up by at least one reviewer. (a) "...be not such a conglomeration of mathematical ideas involved in the scale". (b) "...be composed of short sentences that could be followed verbatim as they are being read by the tester". (c) "...be significantly different from the type of 'word problems' found in most standardized tests." (d) "...include extraneous information", or (alternately) exclude crucial information in which case the required task would be to identify the gap and/or supply a reasonable bit of information to complete a statement of the problem.

Suggested Sets of Word Problems for MANS/77. In response to the above, two examples each of five distinct types of word problems were produced for review by the panelists, writers and coordinators. All the problems were to appear in short sentence form in the student booklet and be read by the tester.

Type I: Missing digits
Type II: Order Relations
Type III: Upper and Lower Bounds
Type IV: Extraneous Data (with two different response formats).
Type V: Two Stage

Time pressure dictated that before reactions could be received from reviewers, each of the above five types be expanded into a 4 or 5 item scale and pilot tested. This was done in three distinctly different 3rd grade classes (selected from the group of six classes involved in the second round of pilot testing). The only planned exception to that procedure was that two alternative scales were made for type IV and each was tested with two classes.

Further discussion of the word problem scales will be given in the scale-by-scale histories which follow.

The 25 scales have been separated into three categories. (1) The 14 scales that made the final version (pages 31-64). (2) The 3 scales that died immediately after being commented upon by the reviewers (pages 65-68). (3) The remaining 8 scales that died somewhere along the way (pages 69-83).

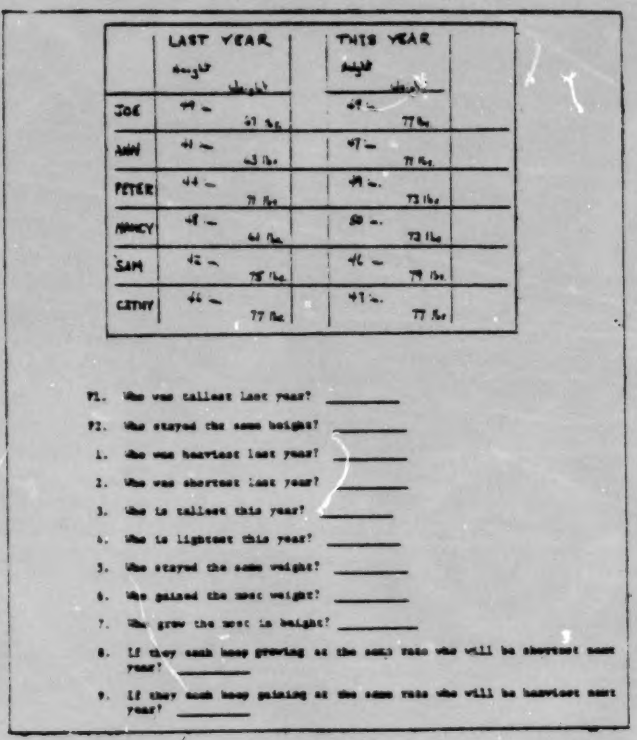
Scales Which Survived

Height and Weight Table
Estimation
Functions
Two Stage Word Problems
*Large Number Computations
Number Line Labelling
Hints and Problems
2 or 5 or 10
Composite Functions
Equation Fluency
Circle the Larger
Missing Digits
Word Problems "With Rounding"
*Fractions

* These two scales should not be considered as regular MANS Scales. The rationale for their inclusion was that many of the MANS/76 scales had fractions, or large number computations, of one kind or another and it was better to group them into homogeneous scales and remove them from other scales intended to test different abilities. They are not novel situations and are more "straight" computation than other scales.

HEIGHT AND WEIGHT TABLE

Initial Version (seen by the reviewers)



	LAST YEAR		THIS YEAR	
	height	weight	height	weight
JOE	49 in.	67 lbs.	49 in.	77 lbs.
ANN	41 in.	63 lbs.	47 in.	71 lbs.
PETER	44 in.	71 lbs.	49 in.	73 lbs.
NANCY	48 in.	61 lbs.	50 in.	73 lbs.
SAM	42 in.	75 lbs.	46 in.	79 lbs.
CATHY	46 in.	77 lbs.	47 in.	77 lbs.

P1. Who was tallest last year? _____

P2. Who stayed the same height? _____

1. Who was heaviest last year? _____

2. Who was shortest last year? _____

3. Who is tallest this year? _____

4. Who is lightest this year? _____

5. Who stayed the same weight? _____

6. Who gained the most weight? _____

7. Who grew the most in height? _____

8. If they each keep growing at the same rate who will be shortest next year? _____

9. If they each keep gaining at the same rate who will be heaviest next year? _____

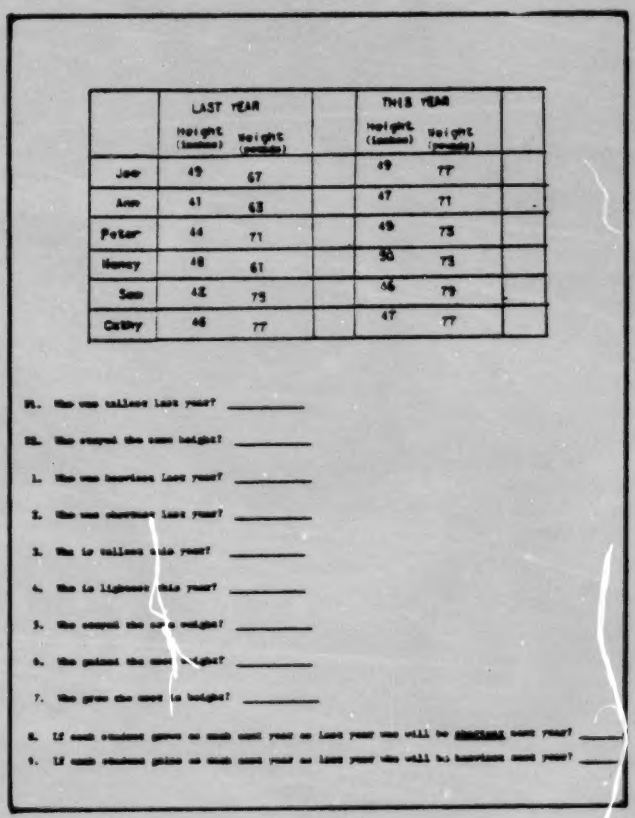
Summary of Instructions: Look at the table for a minute and then we'll do the first two questions together.

Reactions from Writers, Panelists and Coordinators

Three panelists worried about heavy loading on non-mathematical content. One panelist thought the jump in level of difficulty between items 7 and 8 was too great while another thought items 8 and 9 were simply too difficult. One panelist wondered if metric units should be used. Two writers thought the use of the word "rate" in items 8 and 9 was not clear.

Original Versions (seen by the students)

Since some reactions were received before the original version "went to press" for the trials with students, the students received the "benefit" of revised items 8 and 9.



	LAST YEAR		THIS YEAR	
	Height (inches)	Weight (pounds)	Height (inches)	Weight (pounds)
Joe	49	67	49	77
Ann	41	63	47	71
Peter	44	71	49	73
Nancy	48	61	50	73
Sam	42	75	46	79
Cathy	46	77	47	77

P1. Who was tallest last year? _____

P2. Who stayed the same height? _____

1. Who was heaviest last year? _____

2. Who was shortest last year? _____

3. Who is tallest this year? _____

4. Who is lightest this year? _____

5. Who stayed the same weight? _____

6. Who gained the most weight? _____

7. Who grew the most in height? _____

8. If each student grows at each same year as last year who will be shortest next year? _____

9. If each student gains at each same year as last year who will be heaviest next year? _____

Summary of Instructions: Same as above except that the tester first asked orally three simple questions like "How tall was Sam last year?" The answer to each was explained.

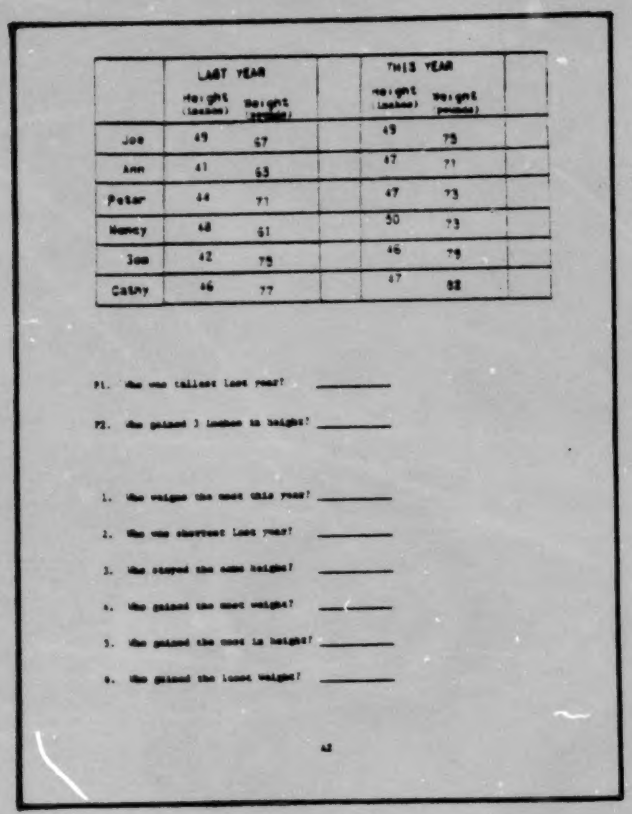
Trials with Students

After a trial with one class it was obvious that students needed help with reading items 8 and 9. Hence for classes two and three, the last two items were read to the students and the instructions were changed accordingly.

In the item analysis on all three classes, the scale checked out quite well except that while the first five items proved to be very easy, the last four proved to be very difficult.

3.3

Hence, in the final version a number of changes were made to try to create a better spread of difficulty level across items. (1) Items 8 and 9 were dropped as being too difficult and impossible to compensate for (2) Example P2 was changed to an item more like the more difficult ones (3) The wording of several items was changed: i.e., previous items 1, 5 and 7 which became new items 1, 6, and 5 respectively (4) Old items 3 and 4 were dropped since new items 1, 2 and 3 were enough at that level of difficulty. As a result of the above changes the table had to be changed in three instances (all "This Year"): Joe's weight, Peter's height and Cathy's weight.



	LAST YEAR		THIS YEAR	
	Height (inches)	Weight (pounds)	Height (inches)	Weight (pounds)
Joe	49	67	49	75
Ann	41	63	47	71
Peter	44	71	47	73
Nancy	48	61	50	73
Jan	42	75	46	79
Cathy	46	77	47	82

P1. Who was tallest last year? _____

P2. Who gained 3 inches in height? _____

1. Who weighs the most this year? _____
2. Who was shortest last year? _____
3. Who stayed the same height? _____
4. Who gained the most weight? _____
5. Who gained the most in height? _____
6. Who gained the least weight? _____

42

Summary of Instructions: Same as above except that an elaborate explanation was prepared for any eventuality in working through P2 with the students.

Initial Version

The original version of this scale was actually a revision of a similar scale from MANS/76 with an important improvement. The MANS/76 version used crazy unequal interval number lines, and received much criticism in that regard. In the MANS/77 version, the line was eliminated.

483 + 337	0	10	50	100	500 X	1000	970 - 532	0	10	100	500	1000
39 + 38	0	10	50	100	500	1000	11 + 11 + 11 + 11	0	10	100	500	1000
101 + 102	0	10	50	100	500	1000	107 - 99	0	10	100	500	1000
479 + 86	0	10	50	100	500	1000	237 + 294	0	10	100	500	1000
173 - 165	0	10	50	100	500	1000	5 x 109	0	10	100	500	1000
53 + 49	0	10	50	100	500	1000	149 - 51	0	10	100	500	1000
19 + 19 + 19	0	10	50	100	500	1000	40 x 10	0	10	100	500	1000
29 x 4	0	10	50	100	500	1000	671 - 662	0	10	100	500	1000

Summary of Instructions: Don't figure out exact answers to the calculations. Not enough time. Just decide between which numbers the answer should be. First two problems are worked out and briefly explained.

Reactions from Writers, Panelists and Coordinators

The scale was universally well received by reviewers. Two serious doubts were expressed by a panelist and a writer. (1) How can one be sure that some students won't calculate the exact answer on the easier items: e.g., 11 + 11 + 11 + 11 or 40 x 10; (2) Again on some items wouldn't the best strategy actually be to calculate rather than estimate: e.g., 53 + 49 or 173-165.

One panelist pointed out that there could be confusion changing from one operation to another, suggesting that the signs of operation be made larger. The same panelist wondered why division was not included. Another panelist thought the word "problem" should be substituted for the word "calculation".

483 + 357	0	10	50	100	500	1000
39 + 38	0	10	50	100	500	1000
185 + 97	0	10	50	100	500	1000
479 + 86	0	10	50	100	500	1000
179 + 165	0	10	50	100	500	1000
56 + 49	0	10	50	100	500	1000
19 + 19 + 19	0	10	50	100	500	1000
29 X 4	0	10	50	100	500	1000

970 − 532	0	10	50	100	500	1000
19 + 29	0	10	50	100	500	1000
107 − 9	0	10	50	100	500	1000
257 + 294	0	10	50	100	500	1000
5 X 109	0	10	50	100	500	1000
149 − 51	0	10	50	100	500	1000
40 X 10	0	10	50	100	500	1000
671 − 662	0	10	50	100	500	1000

Summary of Instructions: Essentially the same except that each page was timed separately (60" for the first, 80" for the second) and students were made well aware of the timed aspect.

First Revision

Concerning the doubts of the reviewers, (1) some of the particularly easy examples (like 11 + 11 + 11 + 11) were eliminated, (2) some of those lending themselves as well to calculation as to estimation were changed (53 + 49 to 56 + 49). Signs of operation were made larger. Division was not included due to the newness of the concept in the typical 3rd grade math curriculum. The word "problem" was reserved for referring to the task the student was required to do. The word "calculation" was reserved for referring to the process implied in each item but not to be done. The examples were separated from the regular items (as suggested.)

Trials with Students

After a trial with one pilot class it was found that the timing aspect simply made some students calculate faster! For the second and third classes, the instructions included more reference the lack of necessity for knowing the exact answer to the calculation.

The item analysis showed (a) the average score was a little low but acceptable, (b) the scale homogeneity measure was moderately high, (c) the range of item difficulty was fairly wide and (d) most students finished all or almost all the items. But there appeared to be difficulty with the switching from one operation to another and the fast calculators had not yet been discouraged.

Final Version MANS/A Pages 3,5,7 and 9

For the final version several changes were made. (1) A separate page was devoted to each of the three operations, and the number of subtraction and multiplication examples was thereby increased (2) In order to alleviate the fast calculator problem, (a) the second example was changed from 38 + 39 to 48 + 49 and an estimation strategy was explained (b) calculating was called "the wrong way to do these problems" (c) references to the lack of time were diminished. (3) The amount of time per 8 problems was increased from 80' to 90' in order to allow more students to finish.

Practice

483 - 357 0 10 50 100 500 X 1000

48 - 49 0 10 50 100 500 1000

ADDITION

185 + 97 0 10 50 100 500 1000

24 + 24 0 10 50 100 500 1000

59 + 39 0 10 50 100 500 1000

479 + 86 0 10 50 100 500 1000

279 + 165 0 10 50 100 500 1000

19 + 29 0 10 50 100 500 1000

257 + 294 0 10 50 100 500 1000

19 + 19 + 19 0 10 50 100 500 1000

SUBTRACTION

100 - 93 0 10 50 100 500 1000

149 - 51 0 10 50 100 500 1000

137 - 125 0 10 50 100 500 1000

970 - 459 0 10 50 100 500 1000

260 - 211 0 10 50 100 500 1000

671 - 662 0 10 50 100 500 1000

700 - 205 0 10 50 100 500 1000

651 - 149 0 10 50 100 500 1000

MULTIPLICATION

5 X 109 0 10 50 100 500 1000

2 X 19 0 10 50 100 500 1000

40 X 10 0 10 50 100 500 1000

4 X 23 0 10 50 100 500 1000

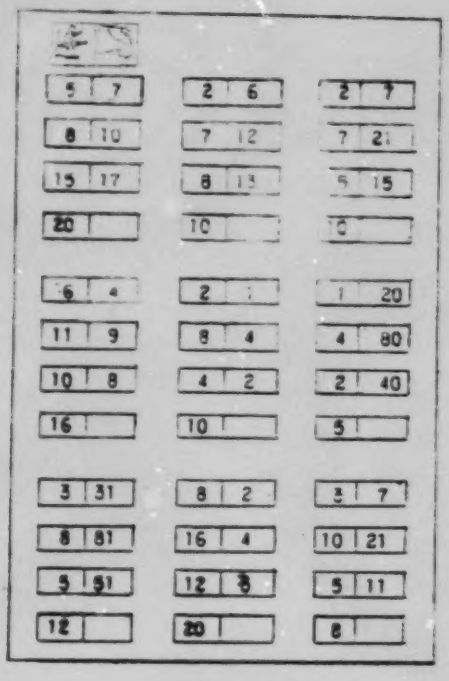
11 X 50 0 10 50 100 500 1000

2 X 49 0 10 50 100 500 1000

4 X 29 0 10 50 100 500 1000

Summary of Instructions: No changes except those indicated on the previous page.

Initial Version



Summary of Instructions: A teacher-led game of guess my rule is explained. The first two rounds are worked through and explained.

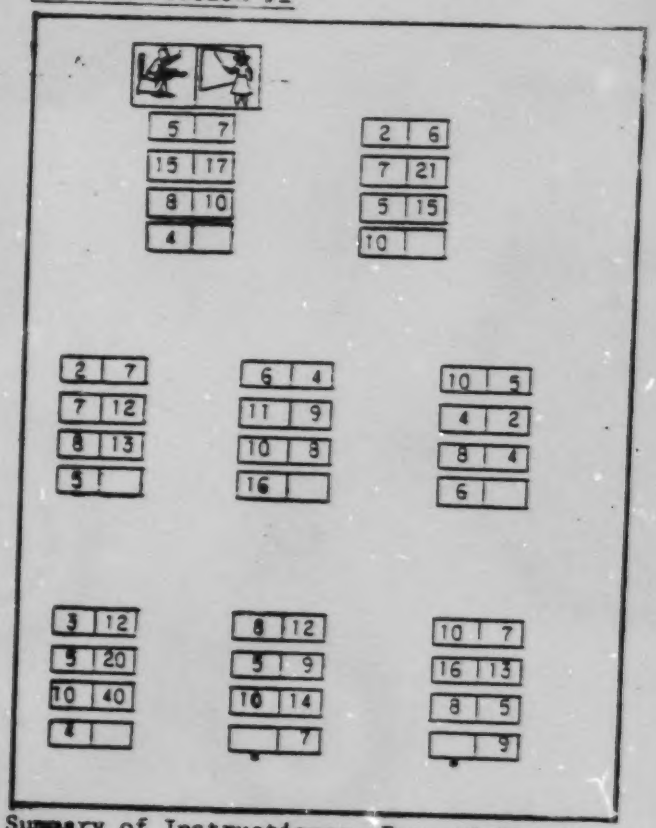
Reactions from Writers, Panelists and Coordinators

Although the reactions were generally positive, one comment was that CSMP students would have a distinct advantage due to their ability to use arrows. Two panelists contradicted this idea by pointing out (a) most math programs in use today have something like this type of activity and (b) the activity seemed more like a non-verbal I.Q. scale than anything else.

Another issue raised by two persons was the matter of guessing. One person thought the activity involved too much guess work, the other thought the insturctions should just reflect the guesswork aspect by saying "guess what goes ...." instead of "figure out what could go . . ."

Other suggested improvements were (a) eliminate rules with two steps, (b) add situations in which the inverse process is expected (c) set examples apart from regular items (d) change format so that student works across rather than down. Suggestions (a) (b) and (c) were heeded in Revised Version #1 and (d) in Revised Version #2

Revised Version #1



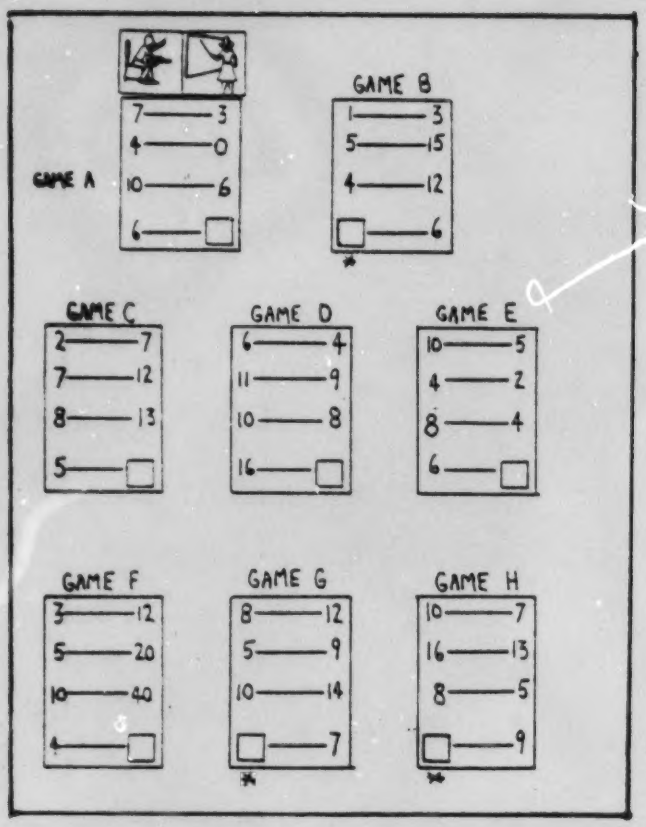
Summary of Instructions: Essentially the same as above.

Trials with Students

In a trial with one class, several problems were encountered: (1) Students tended to just use the original +2 rule every time. (2) They appeared to be working down rather than across. (3) They were very confused by the starred (inverse) items.

Revised Version #2

In reaction to those observations, the 2nd revised version (a) used a subtraction example first (b) used a more vertical format. and (c) used a starred example.



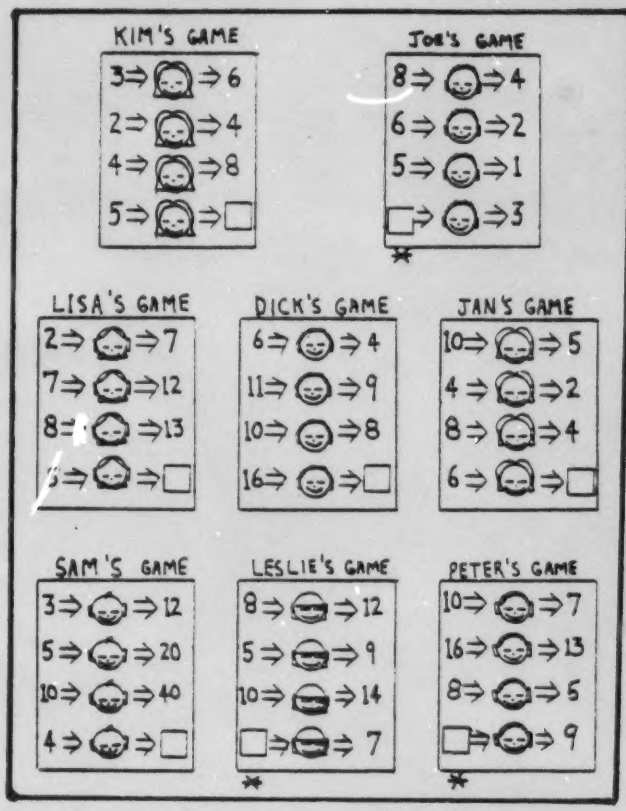
Summary of Instructions: Essentially the same as Revised Version #1

Trials with Students

With two more pilot classes, the item analysis showed that indeed the scale was very difficult: average score 1.8 out of 6.0. Analyses of wrong answers indicated that students were putting the rule in the box rather than the result after using it or simply using the same rule for all items.

Revised Version #3

In an attempt to alleviate the above difficulties, several changes were made. Rather than the "teacher" doing each item, a different student had a different secret rule for each item. In addition the instructions made a point of showing that, when using only the first pair of numbers, there were two highly likely secret rules: (e.g. in example #1, +3 and x2). By using the format which showed the students' faces, it was thought that fewer students would write the rule in the box.



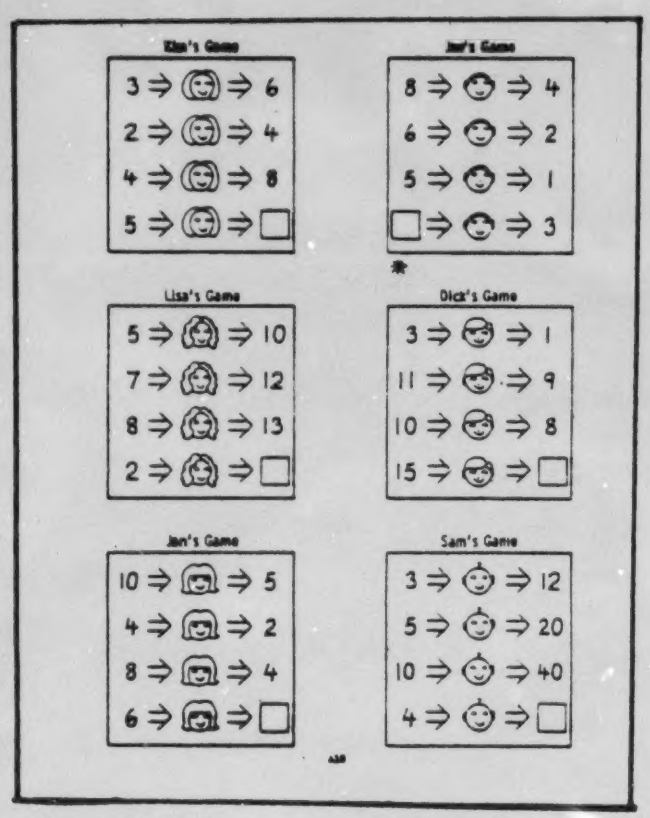
Summary of Instructions: Same except as noted in previous paragraph.

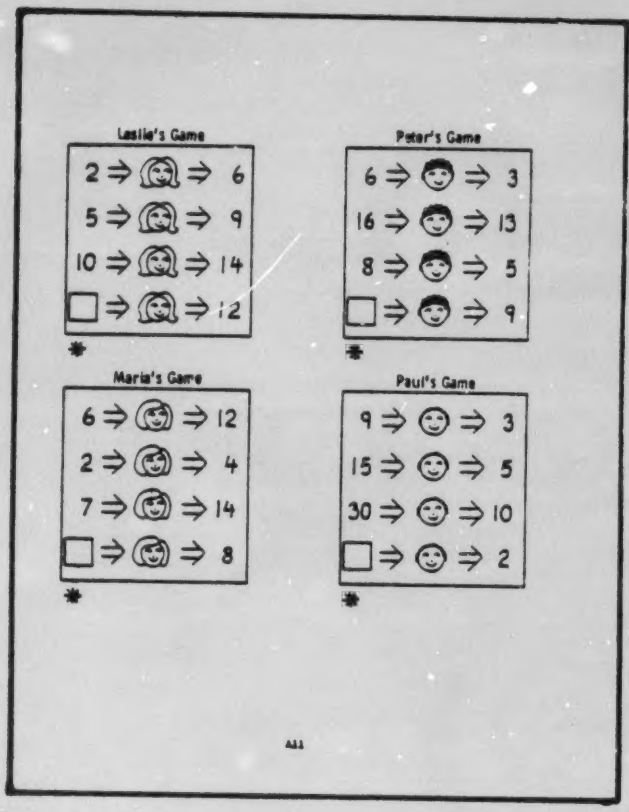
Trials with Students

The third version proved much easier: an average ability class had a mean score of 3.0 out of 6.0; an above average class, 4.8.

Final Version: MANS/A Pages 10 and 11

The only changes made were as a result of more space available. The starred items were expanded from 2 to 4 and were separated graphically to facilitate referring to them in the instructions.





Summary of Instructions: Same as previous.

WORD PROBLEMS: TYPE FIVE: TWO STAGE

Initial Versions

As noted previously each of five word problem scales grew out of a discussion about word problems in general. Before reactions to each scale could be received from the reviewers, a scale had to be developed to take advantage of the second round of pilot testing.

For the Reviewers

A. It is 2800 miles from Los Angeles to New York. It is 3400 miles from New York to Paris. Suppose we go from Los Angeles to New York to Paris, then back to New York, and back to Los Angeles. How many miles would the trip be altogether? Circle the calculation that looks like it answers the question.

A: 3,400 - 2,800 = 600
600 x 2 = [1,200]

B: 3,400 + 2,800 = 6,200
6,200 x 2 = [12,400]

C: 3,400 x 2,800 = [9,520,000]

D: 3,400 + 2,800 = 6,200
6,200 ÷ 2 = [6,202]

B. In 1960, Centerville had a population of 643. Since then, 145 people died or left town and 324 people were born or moved into town. What is the population now? Circle the calculation that looks like it answers the question.

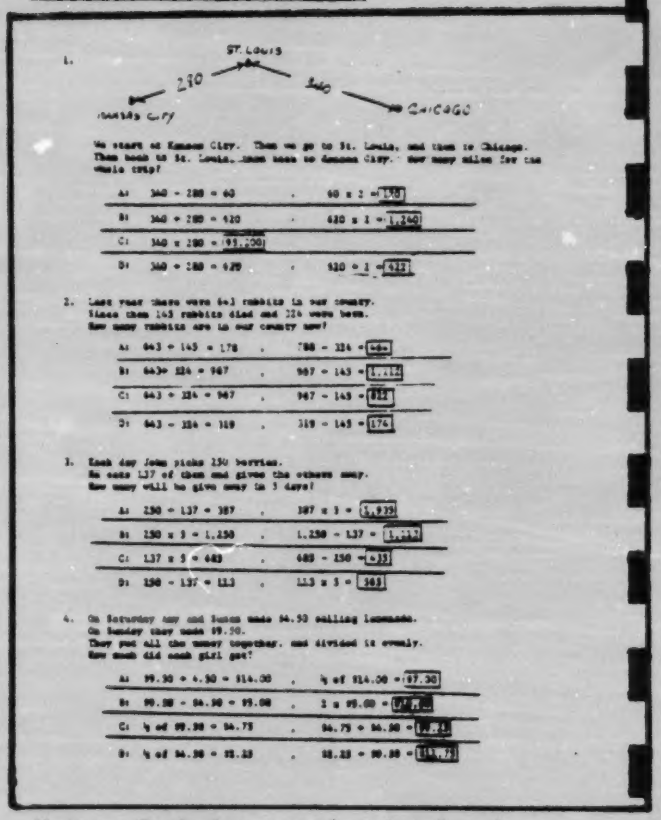
A: 643 + 145
788 - 324 = [464]

B: 643 + 324 = 967
967 + 145 = [1,112]

C: 643 + 324 = 967
967 - 145 = [822]

D: 643 - 324 = 319
319 - 145 = [174]

For the Student Booklet


1.
ST. LOUIS
280
340
KANSAS CITY
CHICAGO
We start at Kansas City. Then we go to St. Louis, and then to Chicago. Then back to St. Louis, then back to Kansas City. How many miles for the whole trip?
A: 340 - 280 = 60 . 60 x 2 = 120
B: 340 + 280 = 620 . 620 x 2 = 1,240
C: 340 x 280 = 95,200
D: 340 + 280 = 620 . 620 ÷ 2 = 310
2. Last year there were 643 rabbits in our county. Since then 145 rabbits died and 324 were born. How many rabbits are in our county now?
A: 643 + 145 = 178 . 788 - 324 = 464
B: 643+ 324 = 967 . 967 + 145 = 1,112
C: 643 + 324 = 967 . 967 - 145 = 822
D: 643 - 324 = 319 . 319 - 145 = 174
3. Each day John picks 250 berries. He eats 137 of them and gives the others away. How many will he give away in 5 days?
A: 250 - 137 = 387 . 387 x 5 = 1,935
B: 250 x 5 = 1,250 . 1,250 - 137 = 1,113
C: 137 x 5 = 685 . 685 - 250 = 435
D: 250 - 137 = 113 . 113 x 5 = 565
4. On Saturday Amy and Susan made $4.50 selling lemonade. On Sunday they made $9.50. They put all the money together, and divided it evenly. How much did each girl get?
A: $9.50 + 4.50 = $14.00 . ½ of $14.00 = $7.00
B: $9.50 - $4.50 = $5.00 . 1 x $5.00 = $5.00
C: ½ of $9.50 = $4.75 . $4.75 + $4.50 = $9.25
D: ½ of $4.50 = $2.25 . $2.25 + $9.50 = $11.75

Note: Each item read twice by the tester. It is pointed out that each calculation has no errors.

Reactions from Reviewers and Pilot Trials

The reviewers' reactions were generally positive. The only significantly negative reaction was by one panelist who thought the items not much different from the usual word problems in standard tests. Several suggestions were made to improve the scale: e.g., use smaller numbers and give students a free response.

From the pilot trials it was discovered that the reading of the items was prohibitively long; fast students went far ahead and had to wait a long time, slower students weren't helped by all that listening. The item analysis showed the average score was 1.8 out of 4.0.

Final Version: MANS/A Page 12

Due to the generally favorable reaction from reviewers and the acceptable item analysis, the scale was selected for MANS/77 with the following changes: (1) The numbers involved in the items were made much smaller and the problem situations changed accordingly, (2) The choice format for the response was eliminated and (3) The scale was lengthened from 4 to 5 items.

1. Last year there were 25 rats on our island.
   Since then 5 rats have died and 3 were born.
   How many rats are on our island now? _____

2. On Saturday Amy and Susan made $13 selling lemonade.
   On Sunday they made $5.
   They put their money together and divided it evenly.
   How much did each girl get? _____

3. Our hens lay 9 eggs every day.
   Each day we eat 6 of them and give the others away.
   During the next 5 days how many eggs will we give away? _____

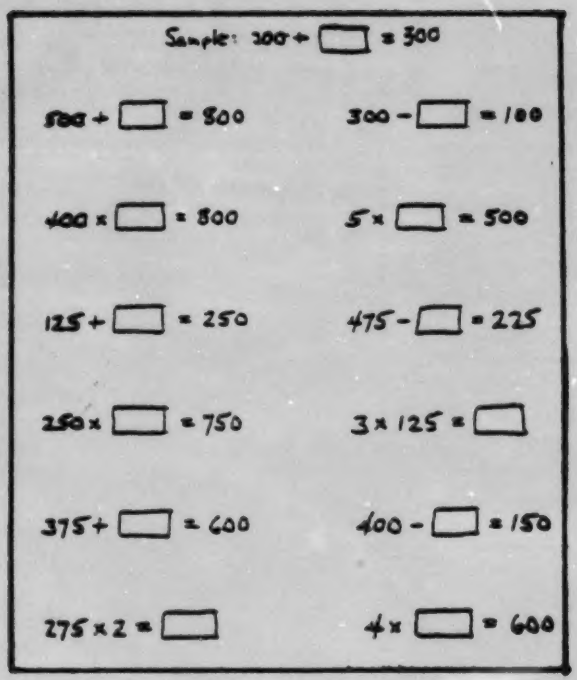
4. Jim has $10 in his bank now.
   Each week he will save $5.
   In how many weeks will he be able to buy a radio that costs $30? _____

5. There are 40 apples in our barrel now.
   We will eat 2 apples every day.
   How many apples will be left in our barrel after 5 days? _____

A12

Note: Each item was read twice by the tester as the student followed along.

Initial Version



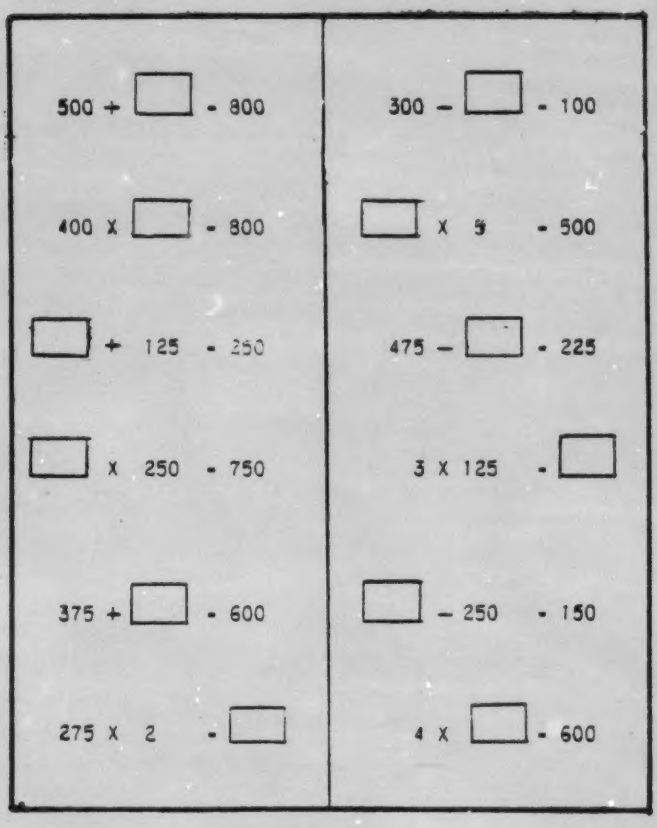
Summary of Instructions: You figure out what goes in the boxes to make the number sentences true. Use empty space for figuring. Some are very hard: skip them if you get stuck. (Went through the sample items together).

Reactions from Writers, Panelists and Coordinators

The scale was acceptable to all eleven reviewers. One panelist suggested including items with the box to the extreme left.

Revised Version

The revised version contains the exact same items as in the initial version except that four items were changed to have the box on the left. The sample item was eliminated as unnecessary.



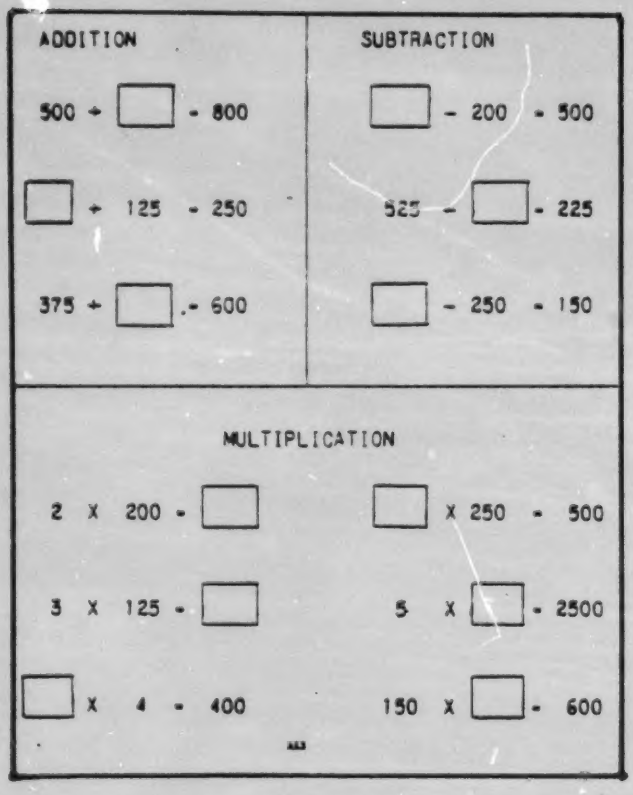
Summary of Instructions: Same as above except for elimination of sample item.

Trials with Students

With three classes, the analysis of wrong answers showed that many students were fooled by the abrupt changes in sign of operation. The item analysis showed that there were no items of middle range difficulty. The mean score was 3.3 out of 12.

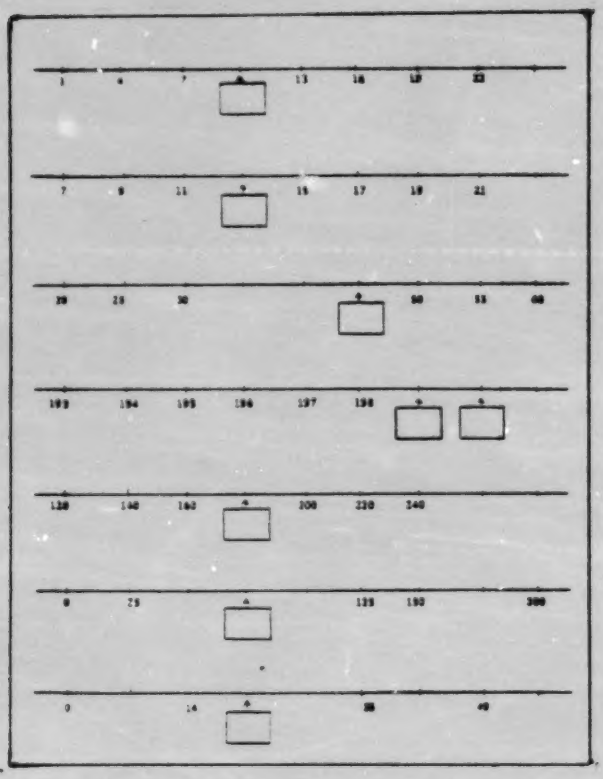
Final Version MANS/A Page 13

The items were separated and labeled according to operation. Seven items were changed slightly in order to make them somewhat easier.



Summary of Instructions: Same as Revised Version.

Initial Version



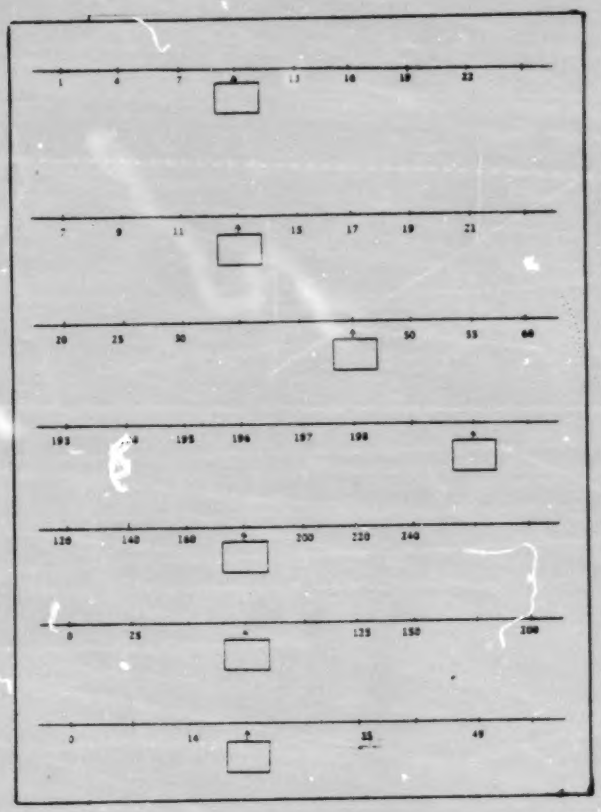
Summary of Instructions: First item is done as a group and explained by the tester.

Reactions from Writers, Panelists and Coordinators

The reviewers' reactions were universally positive. A few improvements were suggested: visually separating the example from the other items, putting only one answer box per line, making the level of difficulty not jump so suddenly with the last item, making the scale either a power or a speed test. One coordinator thought the scale a bit difficult for third graders.

Revised Version

The first two suggestions were heeded and the resultant scale was as follows:



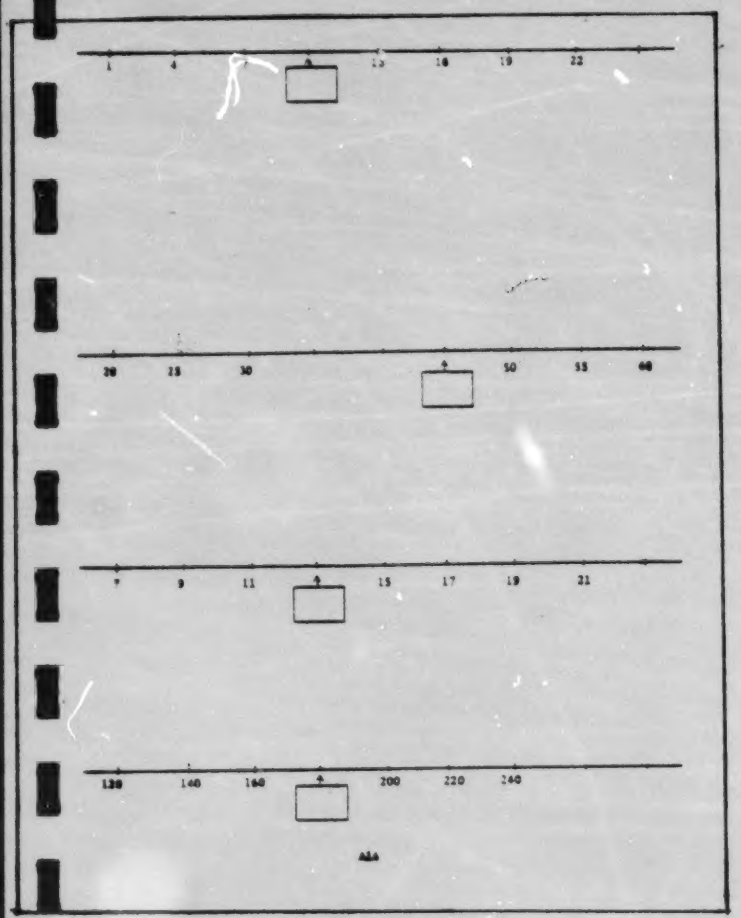
Summary of Instructions: Same as for Initial Version.

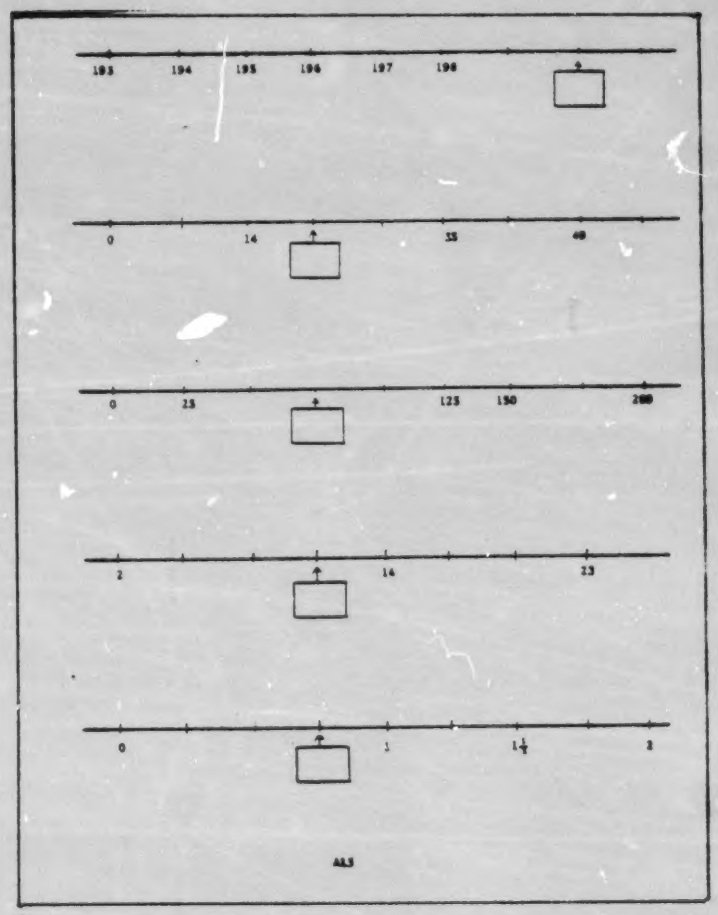
Trials with Students

The trial sessions went very smoothly. The item analysis was satisfactory, the average score being 3.3 of 6.0.

Final Version: MANS/A Pages 14,15

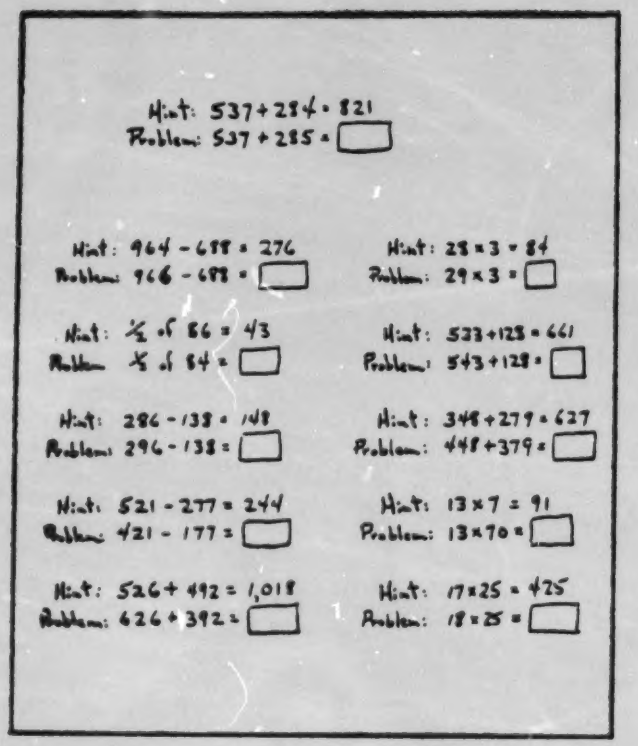
It was decided to lengthen the scale slightly and make it more difficult. The result is shown on next page.





Summary of Instructions: Same as for Initial Version

Initial Version



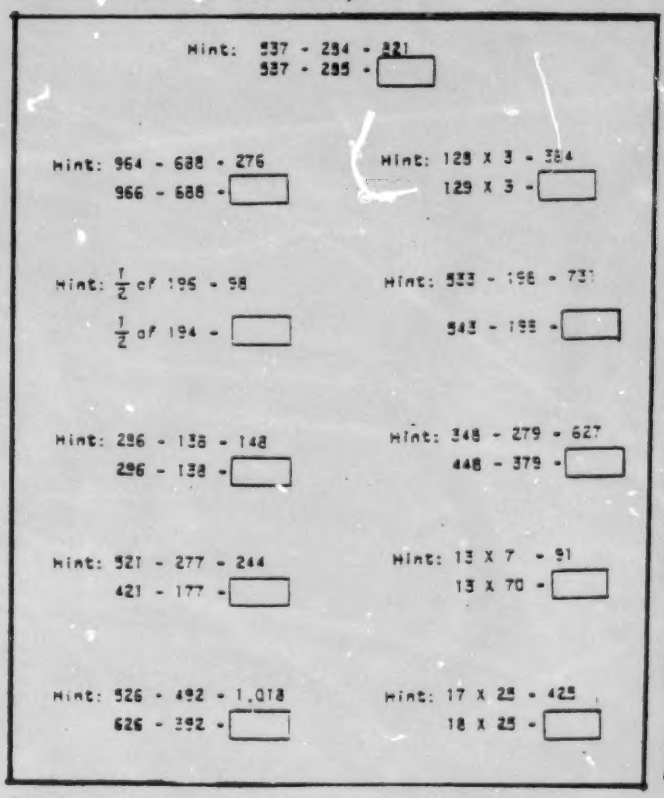
Summary of Instructions: It would take too long to do the calculations. Look at the hint which should make the problem very easy. (Tester then tells them how to do the example: the 285 is one larger than the 284 so .....)

Reactions from Writers, Panelists and Coordinators

The reviewers' reactions were universally positive, but many had suggestions for changes: (a) better not to mix operations on same scale, (b) don't hurry the students in the instructions, just tell them not to do the calculations, (c) example too easy compared to the harder items later, (e) ½ of 84 easier to mentally calculate from scratch, (f) looks visually harder than it actually is, and (g) there must be a better word than "hint".

First Revised Version

The word "problem" was dropped for visual clarity. Three items were changed to make them hard to calculate directly (i.e., without use of the hint).



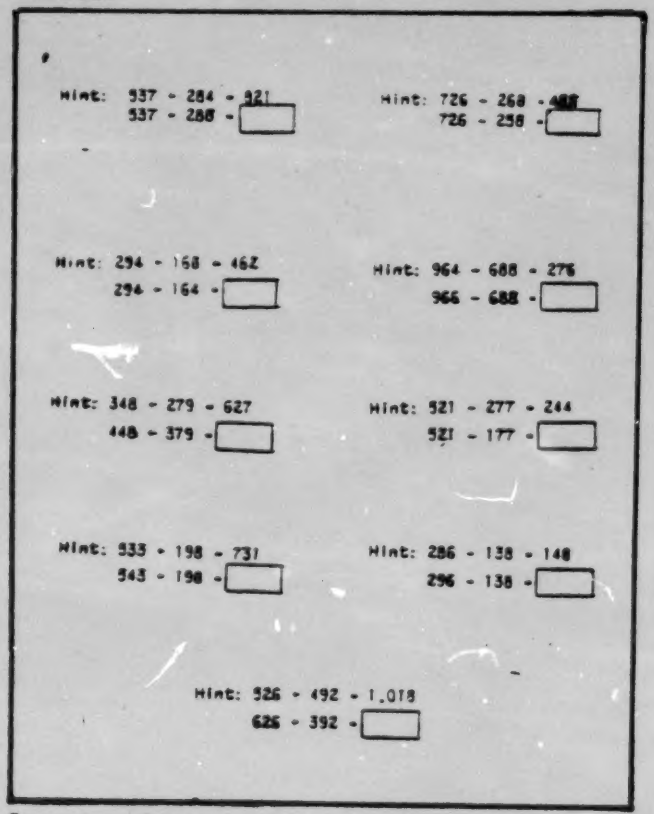
Summary of Instructions: Much less emphasis on time limits, more emphasis on only using the hint. Otherwise, same as initial version.

Trials with Students

After one trial with a class of students it was obvious that a number of changes were called for. In general the scale proved much too difficult for the students, particularly in comparison to the example.

Second Revised Version

The examples were increased from one to two, and both were harder. The number of items were decreased from 10 to 7 and all multiplication items were eliminated.



Summary of Instructions: Essentially same as above except for two examples being explained rather than one.

Trials with Students

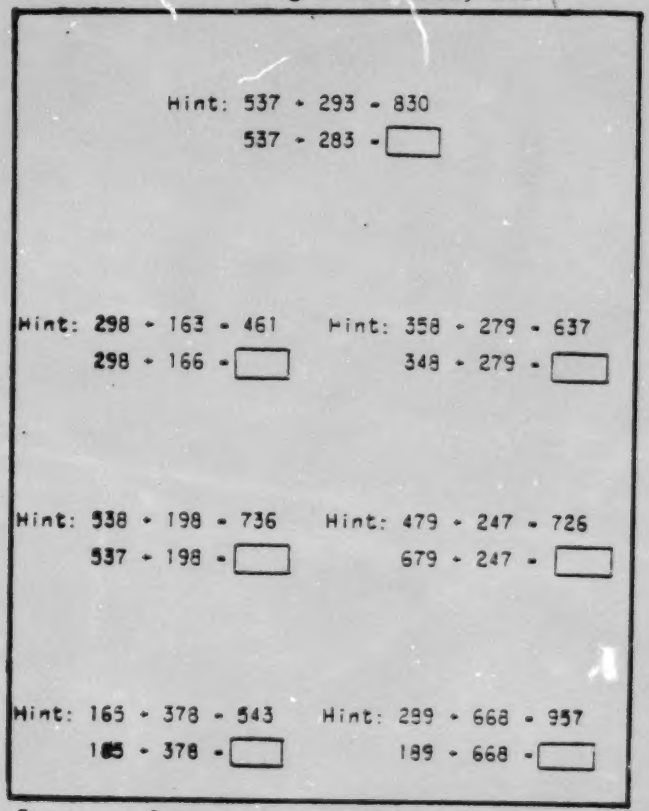
With three additional classes, the item analysis confirmed suspicions that the scale was too difiicult. Average ability classes had average score of 0.9 out of possible 7.0; high ability class, 1.4 out of 7.0. Homogeneity of the scale measured quite low indicating that many students were reduced to just guessing or just increasing the answer to the hint.

In addition, because of the two examples the instructions became rather lengthy for the students.

Third Revised Version

To make the scale easier, the following changes were made: (a) all items were made addition problems, (b) no item involved more than 3-digit numbers, (c) number of items reduced from 7 to 6.

To shorten the instructions and make them less misleading, only one example was used but it involved reducing the answer by ten, followed by the first item which involved an increase of three, followed by an item involving a decrease, etc.



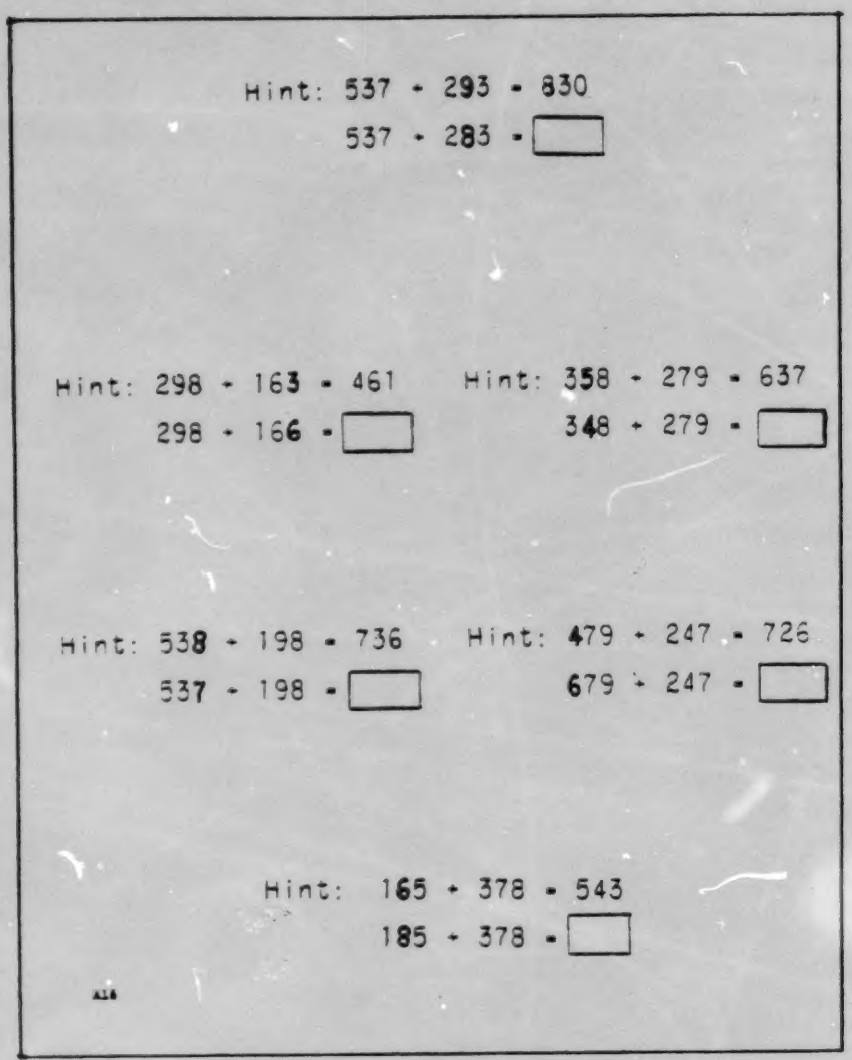
Summary of Instructions: Same as previous, only one example.

Trials with Students

With two classes of average ability the average score was 2.2 out of a possible 6.0: still a bit low, but tolerable. The measure of homogeneity of the scale was very high. No item was missed by more than 3/4 of the students.

Final Version: MANS/A Page 16

The number of items was reduced from 6 to 5. In a final attempt to make the scale yet less difficult, the crucial digit was darkened in the hint and the problem.



Summary of Instructions: Same as previous

5.

Initial Version

2 or 5 or 10

60 is about _____ times as large as 31

51 is about _____ times as large as 5

16 is about _____ times as large as 7

65 is about _____ times as large as 12

200 is about _____ times as large as 21

1C0 is about _____ times as large as 19

98 is about _____ times as large as 50

45 is about _____ times as large as 8

600 is about _____ times as large as 293

75 is about _____ times as large as 8

41 is about _____ times as large as 19

Summary of Instructions: First item is done together and the answer explained by the tester.

Reactions from Writers, Panelists and Coordinators

The reviewers' reactions were universally positive. One panelist thought there should be more items like the last one where the optimum strategy would be to modify both numbers. One writer thought the wording of the items a bit awkward.

Revised Version Number One

Of the last three items, the first two were changed as suggested by the panelist. The last item was changed from 19 to 31 so as not to repeat the 19 from the 6th item. The example was visually separated from the scale itself.

Revised Version Number One

2 or 5 or 10

60 is about ____ times as large as 31

51 is about ____ times as large as 5

16 is about ____ times as large as 7

65 is about ____ times as large as 12

200 is about ____ times as large as 21

100 is about ____ times as large as 19

98 is about ____ times as large as 50

45 is about ____ times as large as 8

602 is about ____ times as large as 298

503 is about ____ times as large as 49

59 is about ____ times as large as 31

Instructions: Same as Initial Version.

First Trials with Students

By observation, the testers found that with an open-ended time limit students were doing three calculations for each item: 2x, 5x and 10x, the 2nd number in each item.

At the same time, the testers thought they observed another significant group that was guessing. This latter idea was confirmed in the item analysis where 5 of the items showed a percentage correct which was very close to chance, and where the homogeneity of the scale was rather low.

Revised Version Number Two

It was decided to put the scale through a second round of testing. In order to cut down on the calculating, the instructions were revised as follows: a time limit of 2½ minutes was set and the students warned at 1¼ minutes that half the time was up.

The scale was tested with CSMP students (the first round being all non-CSMP classes). The scale itself was not changed.

Second Trials with Students

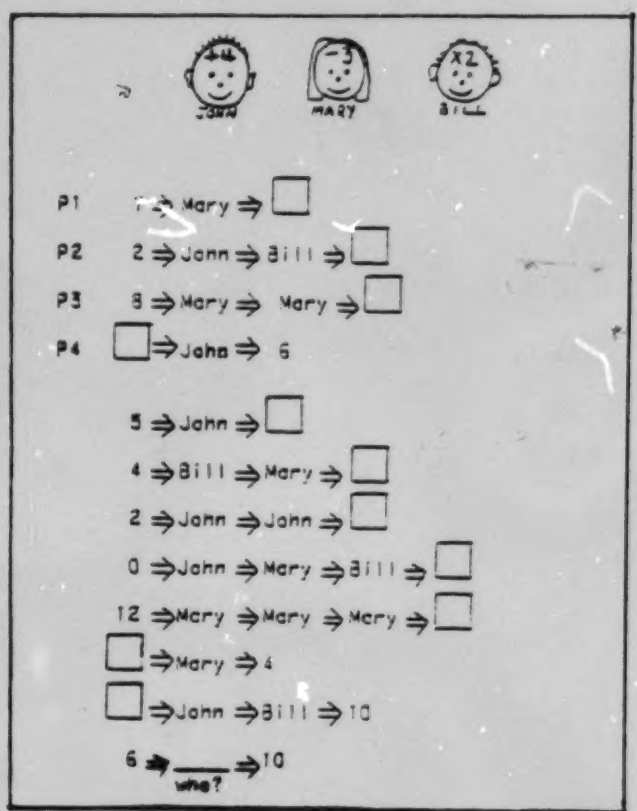
By observation, the testers found less evidence for calculation with the speeded version with CSMP students. In addition the item analysis showed that only 2 of those 5 items (and none others) showed percentages correct that were close to chance. Also the scale homogeneity was satisfactory.

Final Version: MANS/B Page 3

No further revisions were made in the scale as it appeared in the student booklet. The instructions stayed as revised, with a time limit of 2½ minutes and a half time warning.

Initial Versions

This scale grew out of the second round of potential scales and as such was sent to reviewers and pilot tested simultaneously. These two original versions were almost identical except that the first example used 6 (in the reviewer version) and 7 (in the pilot trial version).


JOHN
MARY
BILL
P1 ? ⇒ Mary ⇒ □
P2 2 ⇒ John ⇒ Bill ⇒ □
P3 8 ⇒ Mary ⇒ Mary ⇒ □
P4 □ ⇒ John ⇒ 6
5 ⇒ John ⇒ □
4 ⇒ Bill ⇒ Mary ⇒ □
2 ⇒ John ⇒ John ⇒ □
0 ⇒ John ⇒ Mary ⇒ Bill ⇒ □
12 ⇒ Mary ⇒ Mary ⇒ Mary ⇒ □
□ ⇒ Mary ⇒ 4
□ ⇒ John ⇒ Bill ⇒ 10
6 ⇒ ____ ⇒ 10
who?

Summary of Instructions: John, Mary and Bill are super mental calculator specialists who only perform one operation each, as indicated. Each of the four practice items is read and worked out together: e.g. P2 "...yes 12 is correct. John takes the 2 and adds 4. So he produces 6. Then Bill takes the 6 and multiplies it by 2. That makes 12..."

Reactions from Writers, Panelists and Coordinators

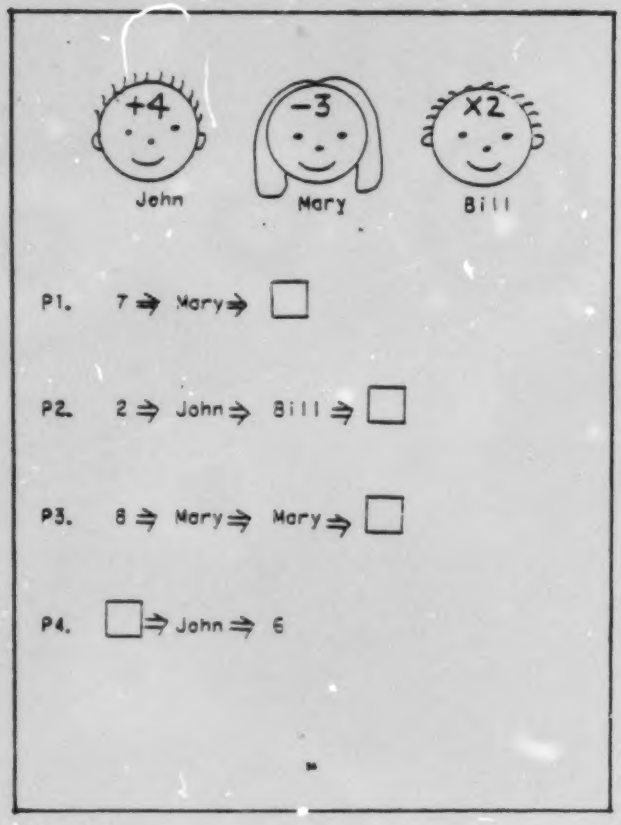
One writer objected to the minimalist definition of "mental calculator" who could do only one thing; the remaining six reviewers found the scale acceptable, with numerous suggested improvements: e.g., use pictures of machines rather than pictures of children's heads, don't use the word "produce", may prove hard to do for all but very brightest.

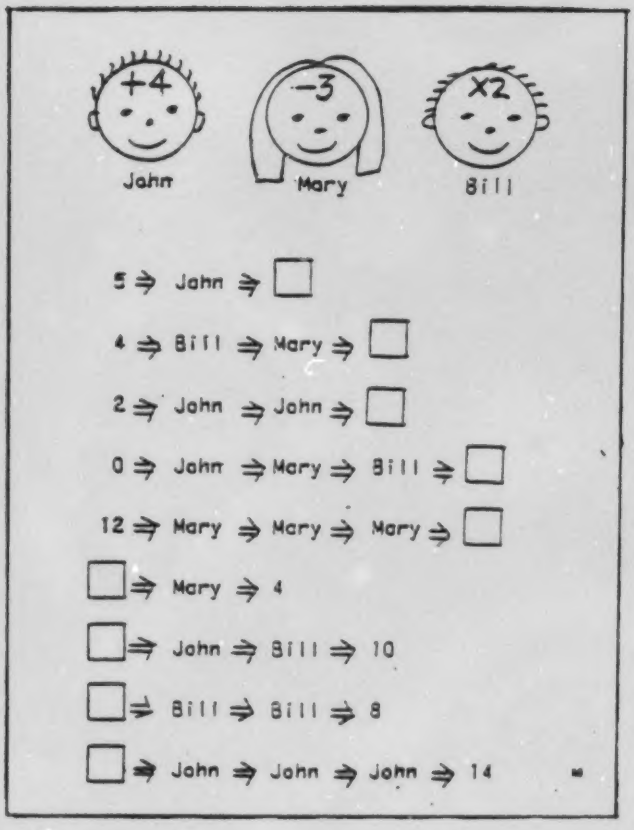
Trials with Students

With three classes it was found that students understood the task (90% got the first item correct and the average score was 4.8 out of 8.0. The scale turned out to be quite homogeneous in spite of essentially three different types of tasks required. A few students were totally confused by the last item.

Final Version MANS/B Pages 4 and 5

The revisions in the final version were (1) the elimination from the instructions of the phrase "mental calculators" (the children had special calculators instead). (2) the addition of two harder items of the second type (3) the elimination of the last (third type of) problem. Since space was available in the booklet two pages were devoted to the scale.




+4
John
-3
Mary
x2
Bill
5 ⇒ John ⇒ ☐
4 ⇒ Bill ⇒ Mary ⇒ ☐
2 ⇒ John ⇒ John ⇒ ☐
0 ⇒ John ⇒ Mary ⇒ Bill ⇒ ☐
12 ⇒ Mary ⇒ Mary ⇒ Mary ⇒ ☐
☐ ⇒ Mary ⇒ 4
☐ ⇒ John ⇒ Bill ⇒ 10
☐ ⇒ Bill ⇒ Bill ⇒ 8
☐ ⇒ John ⇒ John ⇒ John ⇒ 14

Summary of Instructions: Essentially the same as above except for (1) eliminating any idea of the children being mental calculators, (2) pointing out that the last four items were like the practice item P4.

Initial Version

= + − X 1 2 3

3 − 1 = 2

1 X 3 = 3

1 + 1 + 1 + 1 = 2 + 2

Summary of Instructions: Look at the symbols at the top of the page (the signs and numbers were described to the students). We want to write as many different true number sentences as we can using only these signs and numbers. You can use each sign or number as many times as you want in the same sentence. And you don't have to use them all in the same sentence. (The three sentences given as examples were explained.) "Now, I want you to write as many different number sentences as you can."

Reactions from Writers, Panelists and Coordinators

The reviewers' reactions were mixed. One panelist thought that although CSMP considers this type of activity important, he could see little relevance to the "real world of mathematics". One writer had found that classes of students generally had difficulty coping with this type of activity. A panelist and a writer each mentioned the necessity of including parantheses. One panelist thought the inclusion of the number one, made the exercise "trivial". Another panelist pointed out the problem of power versus speed. But half the reviewers found the scale entirely acceptable.

Revised Version No. 1

− + − X 1 2 3 ( )

3 − 1 = 2

1 X 3 = 3

1 + 1 + 1 + 1 = 2 X 2

2 X (3-1) = 2 + 2

Summary of Instructions: Same as Initial Version except that it was emphasized that only the numbers one, two and three could be used. The parantheses were mentioned and the example was given, but no futher explanation was made of their use.

Trials with Students

In a trial with one class it was apparent that further revision was necessary. The use of numbers other than 1, 2 and 3 was widespread. Students were confused by the presence of parentheses.

Revised Version No. 2

	+	1	
-	-	2	( )
	X	3	

3 - 1 = 2	
1 X 3 = 3	
1 + 1 + 1 + 1 = 2 X 2	
2 X (3-1) = 2 + 2	

Summary of Instructions: Same as above except that the symbols were introduced in groups (giving further emphases to 1, 2, and 3 being the only allowable symbols. In addition it was pointed out that parentheses didn't have to be used if one didn't want to.

Trials with Students

There was still wide-spread use of numbers other than 1, 2 and 3 particularly on the right side of the equation sign. It was also observed that some students wrote equations only up to the point that lines were available.

Revised Version No. 3

	+	1	
-	-	2	( )
	X	3	

3 - 1 = 2	
2 X 3 = 6	
1 + 1 + 1 + 1 = 2 X 2	
2 X (3-1) = 2 + 2	

Summary of Instructions: To emphasize further the incorrect use of numbers other than 1, 2 and 3, the second example was purposely printed incorrectly. Students were asked what was wrong with it, then told to cross it out and write 1 x 3 = 3. Students were asked to circle their last equation after 2, 4 and 6 minutes and given 8 minutes altogether.

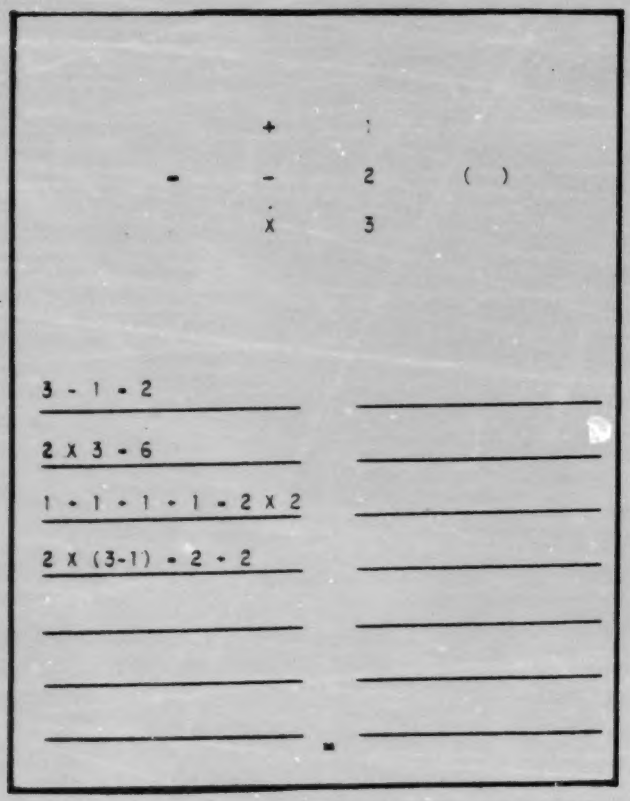
Trials with Students

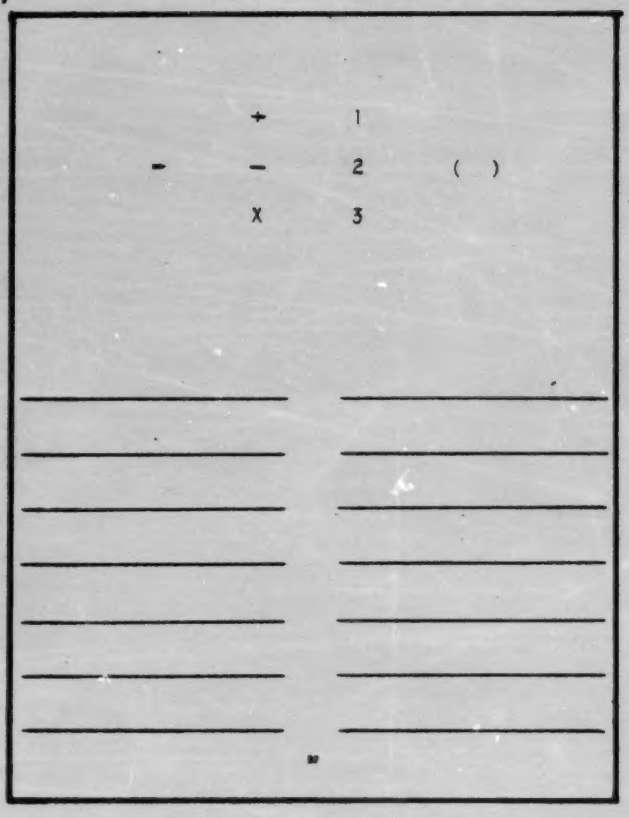
With four classes the use of numbers other than 1, 2 or 3 was virtually eliminated. Very few students wrote equations after four minutes. Some students indeed stopped writing after running out of available lines.

5.

Final Version: MANS/B Pages 6 and 7

In the final version, the number of lines was increased form 16 to 24, putting the scale on two pages. The instructions were not changed except that the students were given 5 minutes with no circling after 2 or 4 minutes.





Initial Version

SAMPLE :	5 x 9	7 x 9
	53 + 19	53 + 23
	183 - 156	193 - 156
	$\frac{1}{4}$ of 346	$\frac{1}{5}$ of 346
	7 x 241	7 x 341
	424 + 388	423 + 390
	652 - 180	652 - 190
	6 x 163	7 x 160
	$\frac{1}{6}$ of 372	$\frac{1}{5}$ of 372
	450 - 178	460 - 198
	47 + 67 - 27	57 + 77 - 37
	2 x 51	$\frac{1}{7}$ of 198
	25 - 103 + 25	30 - 103 + 30
	77,777,777	77,687,777
	72 + 72 + 72 + 72	72 x 5
	$\frac{1}{5}$ x 812	$\frac{1}{3}$ x 330

Summary of Instructions: For each pair of calculations, put an "x" on the larger of the two. Don't do the calculations, just estimate or you'll run out of time.

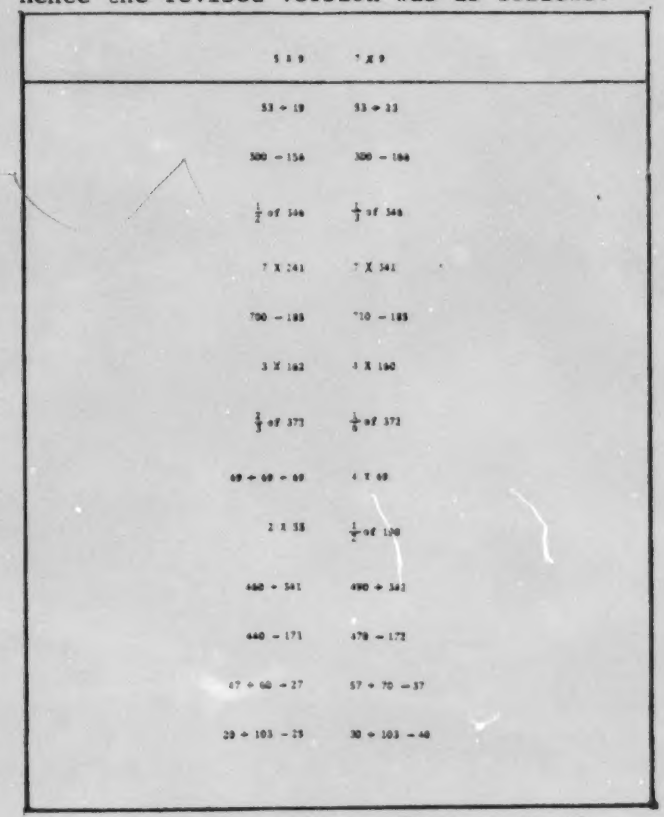
Reactions from Writers, Panelists and Coordinators

Except for one panelist (who was really bothered by it) the scale was universally well received. There were however a number of suggested improvements: make the required task more homogeneous, "circle" instead of "x" the larger, some items are too difficult conceptually or would have to be calculated out, what about fast calculators versus good estimators, need many items to discriminate above the guessing scores, be careful with right versus left on correct responses, use a format requiring students to place "=", ">", or "<" between each pair of calculations.

Revised Version:

In the revised version an attempt was made to heed each of the suggested improvements except the last one. In that case it was decided that the alternate format would change the scale to a test of symbols "=", " > " and " < " as well as a test of estimation ability. Of particular note was the warning about fast calculation versus good estimation, since the scale was supposed to measure the latter, not the former. To partially overcome this problem, the scale was speeded, that is, a short time was allowed. Students were not to calculate, and were alerted when half the time was up. (As for the "too many items required for above chance" the item analysis was used to measure the seriousness of this problem).

Hence the revised version was as follows:



Summary of Instructions: Same as previous except "circle the bigger" rather than "x" it. A demonstration card was used to illustrate correct answering procedure on the example. The word "of" was defined as "times". Two and a half minutes were given with warning at 1¼ minutes.

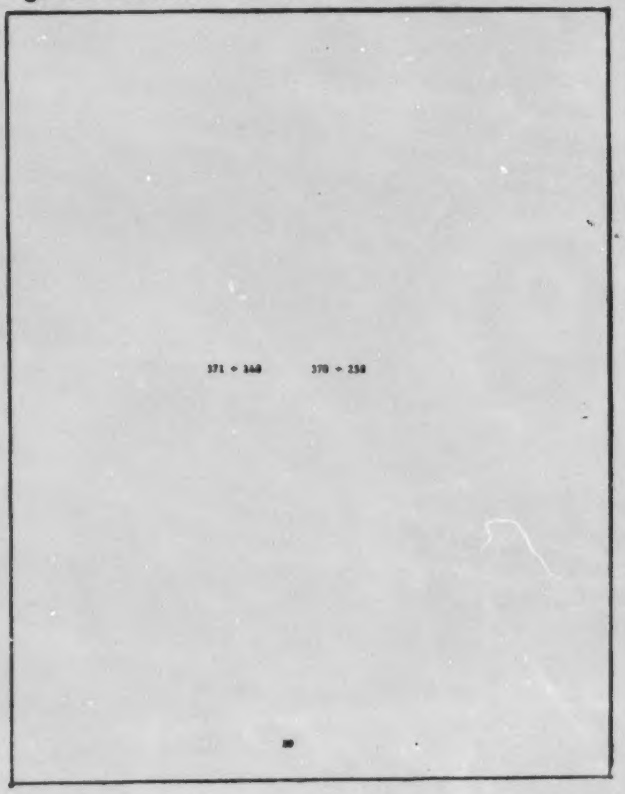
Trials with Students

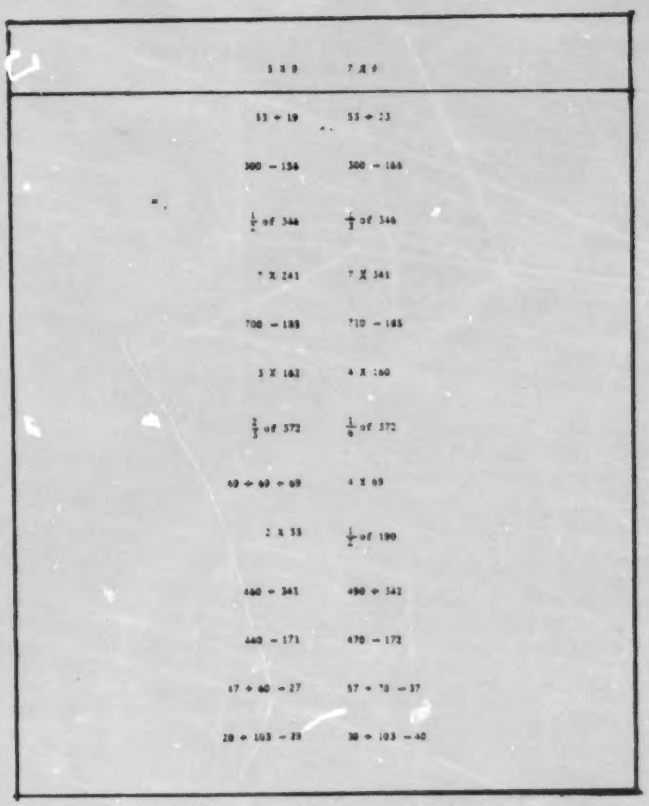
The revised version was tested with 3 classes of students. Students had little difficulty understanding the task required. The item analysis gave an average score of 6.9 on the 13 point scale. Eighty-four percent of the students completed the scale. A bimodal distribution of item difficulty reaffirmed the observations that few students were merely guessing, but indicated the need for more items of middle level difficulty.

Final Version: MANS/B Pages 9 & 11

In order to obtain a more normal distribution of item difficulty the hardest items were eliminated and a harder item was used as the example (explained in the instructions).

The resultant scale and instructions are given below:





Summary of Instructions: Same as in the revised version except (a) "larger" was substituted for "bigger", (b) total time allowed was increased to 3 minutes, (c) a more difficult example was used and a more involved explanation was given to justify the correct response: consequently the example was given on a blind page so that quicker students couldn't jump the gun and hence have more time.

"WORD PROBLEMS" TYPE ONE: MISSING DIGITS

Initial Versions

As noted previously each of five word problem scales grew out of a discussion about word problems in general. Before reactions to each scale could be received from the reviewers, a scale had to be developed to take advantage of the second round of pilot testing.

For the Reviewers

A. A scrap of paper is found on the floor with some figuring.

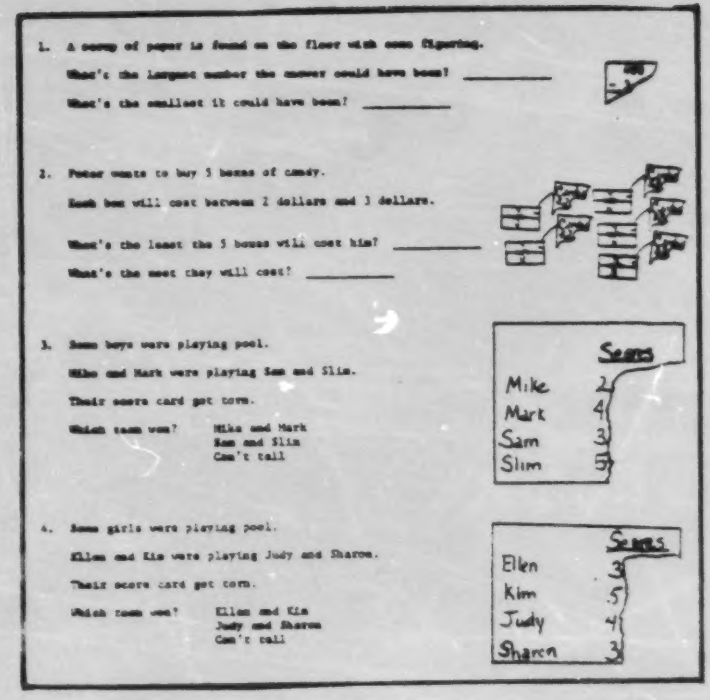
What's the largest number the answer could have been? ____

What's the smallest it could have been? ____

B. Peter wants to buy 5 boxes of candy. He knows that each box will cost more than 2 dollars and less than 3 dollars. What's the least the 5 boxes will cost him? ____

What's the most they will cost? ___

For Student Booklet


1. A scrap of paper is found on the floor with some figuring.
What's the largest number the answer could have been? _____
What's the smallest it could have been? _____
2. Peter wants to buy 5 boxes of candy.
Each box will cost between 2 dollars and 3 dollars.
What's the least the 5 boxes will cost him? _____
What's the most they will cost? _____
3. Some boys were playing pool.
Mike and Mark were playing Sam and Slim.
Their score card got torn.
Which team won? Mike and Mark
Sam and Slim
Can't tell
Scores
Mike 2
Mark 4
Sam 3
Slim 5
4. Some girls were playing pool.
Ellen and Kim were playing Judy and Sharon.
Their score card got torn.
Which team won? Ellen and Kim
Judy and Sharon
Can't tell
Scores
Ellen 3
Kim 5
Judy 4
Sharon 3

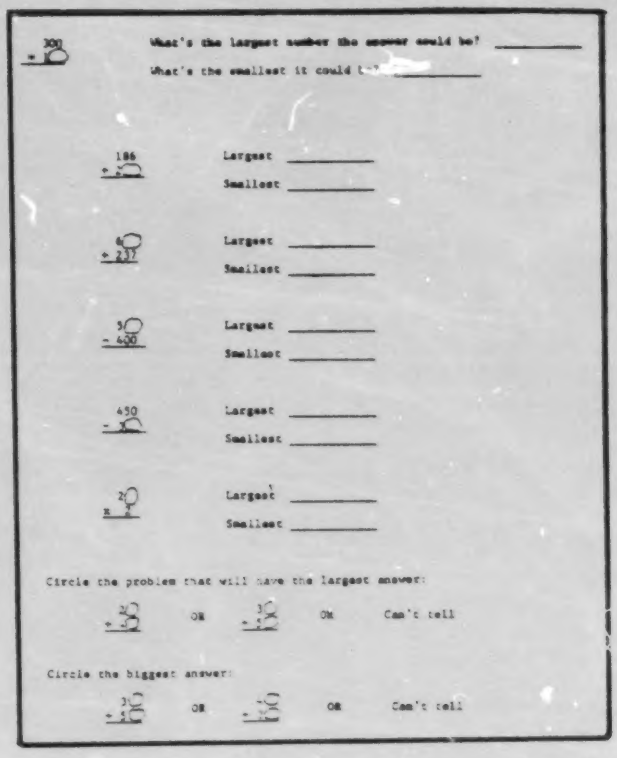
Note: Each problem to be read twice by the tester.

Reactions from the Reviewers and Pilot Trials

The reactions from reviewers were positiv in the main. But one reviewer thought the stories too artificial and time consuming. To another reviewer, the items didn't really seem to be word problems.

The pilot testing went particular badly. The original version of the scale proved to be extremely difficult for the first class with which it was used: average score 1.3 out of a possible 6.0. A completely revised version was produced and used with the other two classes; it is shown below.

Revised Version

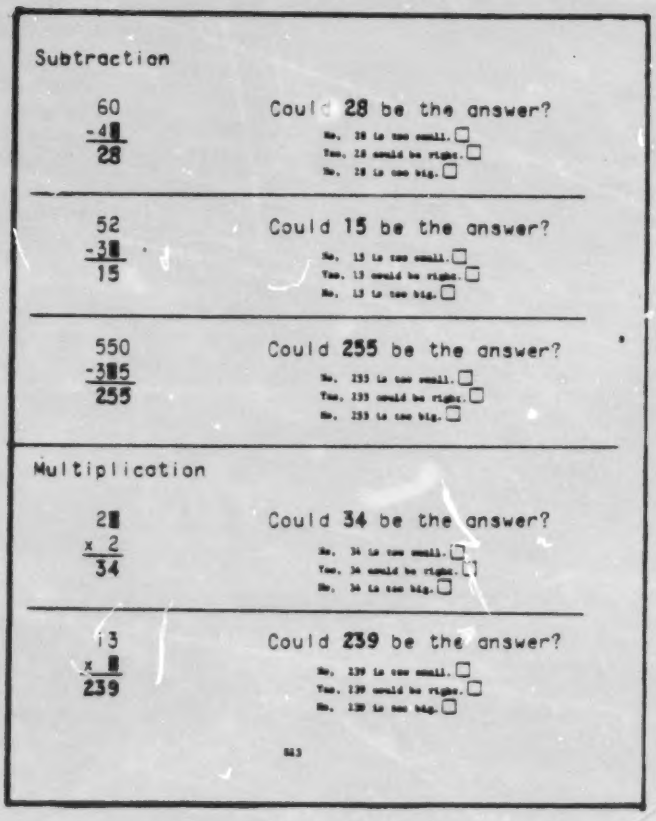


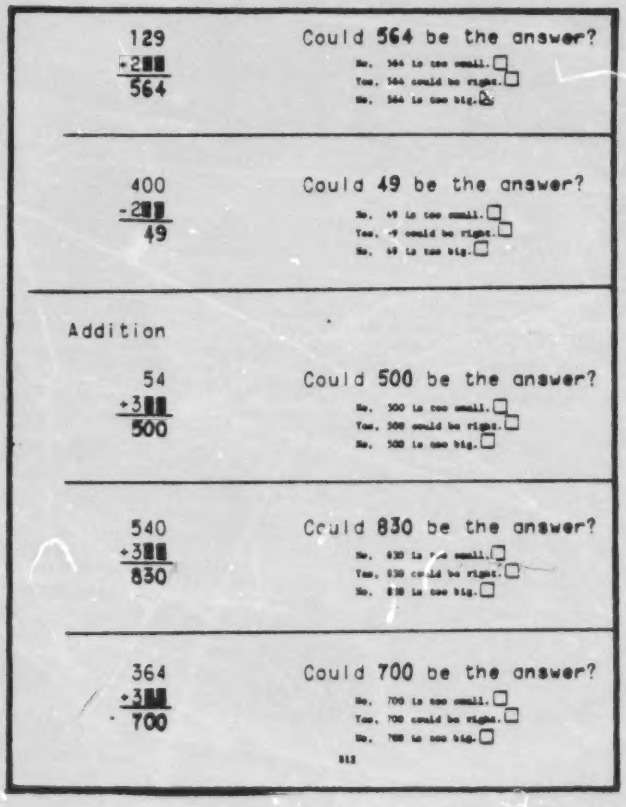
Summary of Instructions: The example at the top was worked through as a group. The O was explained as an ink blotch that covered the digits originally printed there. Attention was called to the two unusual problems at the bottom of the page, particularly the "can't tell" choice.

In the trials with the other two classes, the revised version fared only slightly beteer than the original version did with the first class. Several major problems were still apparent: e.g., average score still very low, 3.1 out of a possible 12. Five of the last 6 items proved extremely difficult.

FINal Version: MANS/B Pages 12 & 13

In spite of continued difficulties with the original and revised versions, it was decided that since the scale was an acceptable one (at least in concept), it would be worthwhile making another revision and including it in the final version of the MANS/77.


Subtraction
60 -4■ 28
Could 28 be the answer?
No, 28 is too small.
Yes, 28 could be right.
No, 28 is too big.
52 -3■ 15
Could 15 be the answer?
No, 15 is too small.
Yes, 15 could be right.
No, 15 is too big.
550 -3■5 255
Could 255 be the answer?
No, 255 is too small.
Yes, 255 could be right.
No, 255 is too big.
Multiplication
2■ x 2 34
Could 34 be the answer?
No, 34 is too small.
Yes, 34 could be right.
No, 34 is too big.
13 x ■ 239
Could 239 be the answer?
No, 239 is too small.
Yes, 239 could be right.
No, 239 is too big.
843


129 +2■■ 564
Could 564 be the answer?
No, 564 is too small.
Yes, 564 could be right.
No, 564 is too big.
400 -2■■ 49
Could 49 be the answer?
No, 49 is too small.
Yes, 49 could be right.
No, 49 is too big.
Addition
54 +3■■ 500
Could 500 be the answer?
No, 500 is too small.
Yes, 500 could be right.
No, 500 is too big.
540 +3■■ 830
Could 830 be the answer?
No, 830 is too small.
Yes, 830 could be right.
No, 830 is too big.
364 +3■■ 700
Could 700 be the answer?
No, 700 is too small.
Yes, 700 could be right.
No, 700 is too big.
813

Summary of Instructions: The calculations on the left were done by a student before some of the digits were covered up. Let's work through the first two problems and find out whether the student's answer is too big, possibly correct, or too small.

WORD PROBLEMS: TYPE THREE; "WITH ROUNDING"

Initial Versions

As noted previously each of five word problem scales grew out of a discussion about word problems in general. Before reactions to each scale could be received from the reviewers, a scale had to be developed to take advantage of the second round of pilot testing.

For the Reviewers

A. Our job is to give three sheets of paper to each person who comes by. We only have fourteen sheets of paper left. For how many more people will we be able to do our job? _____

B. We are making up a comic book. Four pictures will fit on each page. We have 14 pictures to put in the book. How many pages will we need? _____

For the Student Booklets

1. Our job is to put 3 nails in each board.
   We only have 14 nails.
   For how many boards can we do our job? _____

2. We are making up a comic book.
   ~~Four~~ 4 pictures will fit on each page.
   We have 14 pictures to put in the book.
   How many pages will we need? _____

3. A factory runs 18 hours a day.
   The factory paints airplanes.
   It takes 5 hours to paint a whole airplane.
   How many airplanes can the factory paint in one day? _____

4. An elevator only holds 5 people.
   ~~Twenty-three~~ 23 people want to ride to the top.
   How many trips will the elevator need to make? _____

Note: Tester reads each item twice as studen follows along.

Reactions from Reviewers and Pilot Trials

Except for one writer (who found the items trivial and of no interest), the reviewer reactions were generally positive. One panelist thought an initial practice item would be helpful. One writer thought more precise language was called for.

From the pilot trials it was discovered that the items were not easily understood (an abundance of wild wrong answers) and the average score was a low 1.1 out of a possible 4.

Final Version: MANS/B Pages 15

Due to the fact that the scale was suggested by one panelist and received generally positive review, it was decided that it should be included in the MANS/77 with the following revisions (1) the items would be changed so as to sound more realistic to third graders. (2) The number of items was increased from 4 to 5 (3) Students were warned that their answers should not use fractions.

1. An elevator can't hold more than 3 people.
   23 people want to ride to the top floor.
   How many times will the elevator have to go up? _____

2. Mr. Jones owns a car transporting company.
   His trucks can carry 3 cars each.
   He is loading 14 cars onto trucks.
   How many trucks will he need? _____

3. It takes 4 men to lift a piano.
   We have 14 men ready to work.
   How many pianos can they lift at the same time? _____

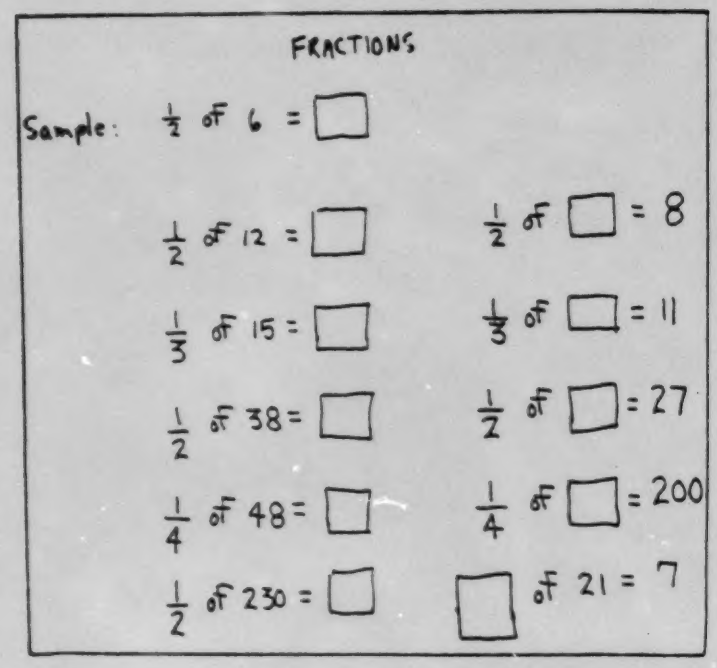
4. There are 18 hamsters in a pet store.
   The store owner is putting the hamsters into new cages.
   Each new cage holds no more than 4 hamsters.
   How many cages will be need? _____

5. A school wants to send as many volleyball teams as possible to a tournament.
   Each team must have 6 players.
   There are 21 volleyball players in the school.
   How many teams can they send? _____

R13

Note: Same as previous

Initial Version

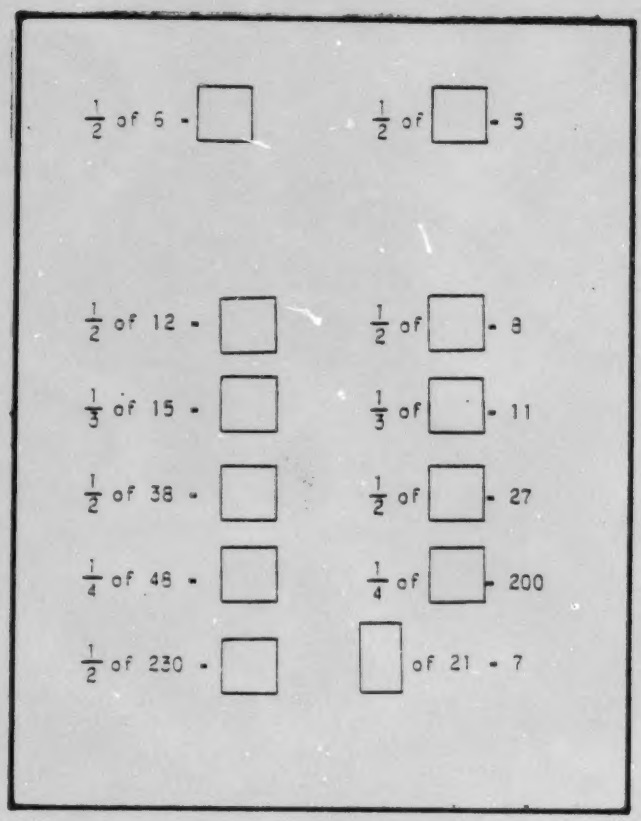


Summary of Instructions: You figure out what goes in the boxes to make the number sentences true. Use empty space for figuring. Some are very hard: skip them if you get stuck. (Went through the sample items together).

Reactions from Writers, Panelists and Coordinators

All reviewers found the scale acceptable. One panelist suggested two samples, the extra one of the type in the right hand column.

Revised Version Number One



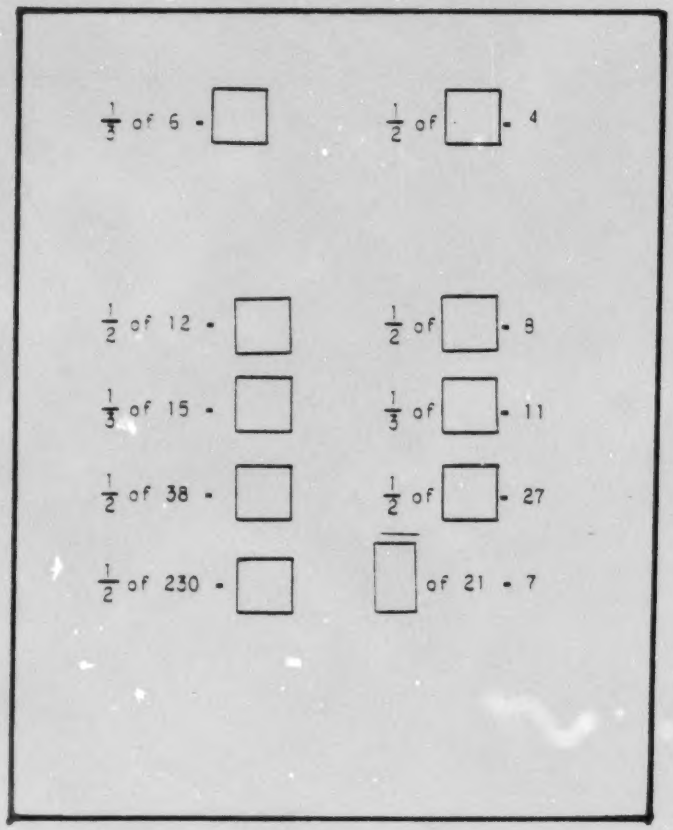
Summary of Instructions: Same as previous except for working through two examples rather than one.

Trials with Students:

With the first class it was obvious that the scale was too difficult, particularly the items involving "¼ of_". In addition it was thought that the second example was too easy a preparation for others of its type.

Revised Version Number Two

The items with "⅓ of _" were eliminated and the second sample was changed to ½ of □ = 4, raising the distinct possibility of the wrong answer "2" and a negative response from the tester.



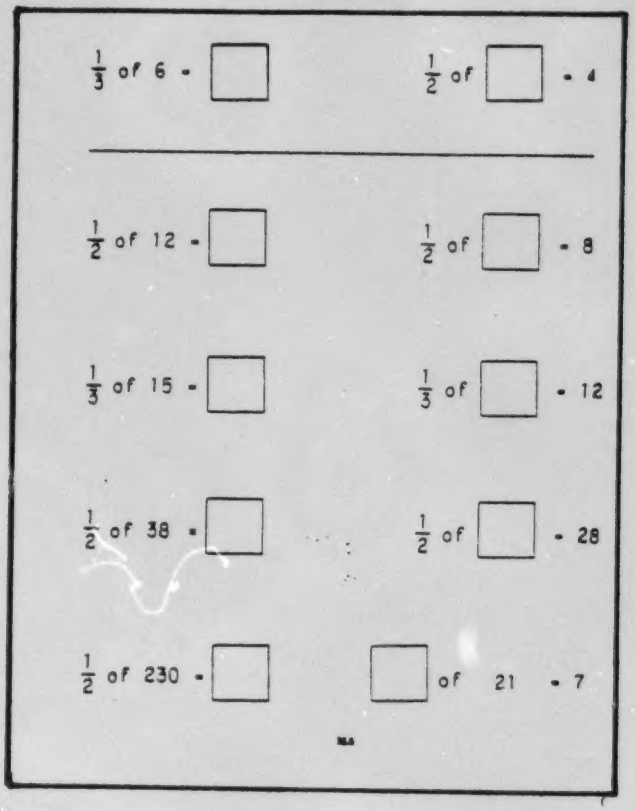
Summary of Instructions: Same as previous

Trials with Students

The item analysis of the first three classes showed the scale to still be too difficult: average score 2.2 of 8.0. This was particularly true of the items in the right hand column. Also, many students gave wrong answers indicating that they didn't have much of a clue as to what was required.

Final Version MANS B Page 16

Since the trials were conducted with non-CSMP students exclusively, (i.e., not expected to do well with items of this type) only a small effort was made to make the scale easier for students to cope with: i.e., items 2 and 3 in the right hand column were changed to be more of the type illustrated in the example.



Summary of Instructions: Same as above except that the second example was discussed in such a way that the wrong answer "2" was always brought up and discounted.

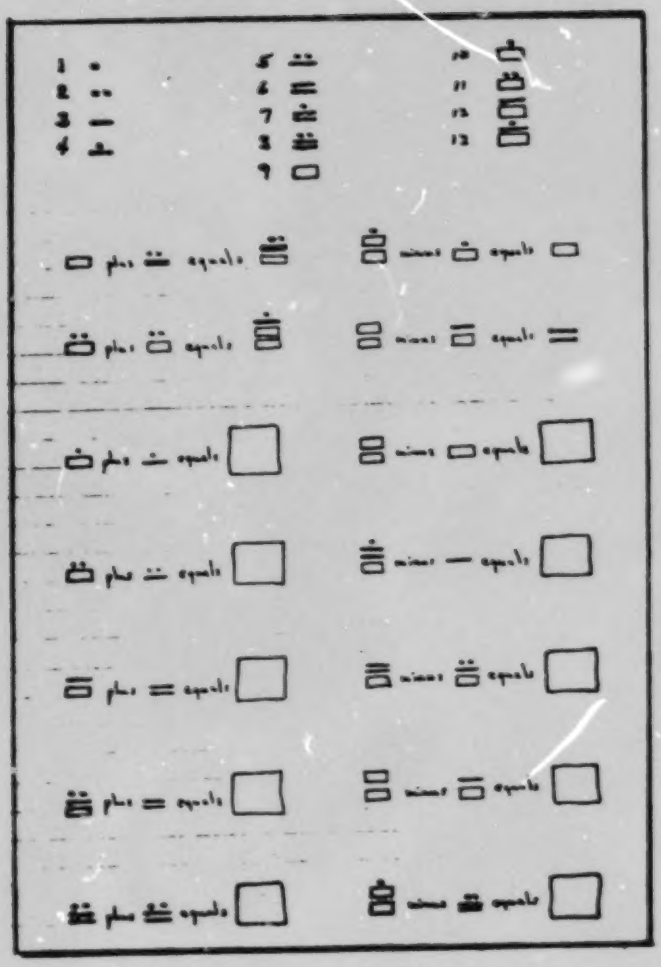
Scales Which Were Immediately Rejected

Goolan Arithmetic

Equations in Two Variables

Dot Pictures

Initial Version



Summary of Instructions: In Goola, the Goolans write numbers with dots (.), bars (-), and blocks (▭). The Goolan writing of the numbers one through thirteen is given at the top. The examples each are given for addition and subtraction problems as Goola children do them. The answers to the ten items are to be given in Goolan.

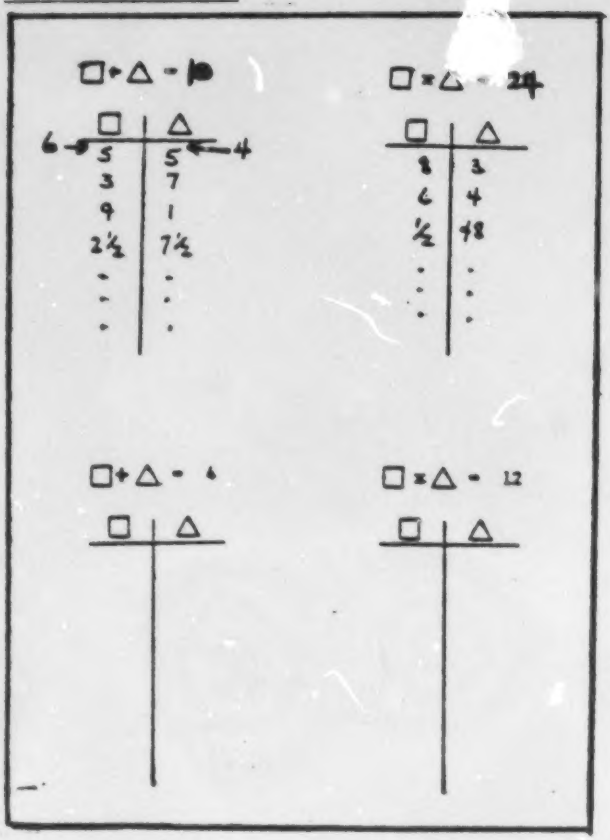
Reactions from Writers, Panelists and Coordinators

Only three of eleven reviewers did not have serious objections to the use of this scale. Some of the substantive reasons given were as follows: too much of an exercise in looking up at the table rather than understanding numberation, maybe a test of intelligence (i.e., coping with a new deciphering situation) more than anything else, looks too much like modern math.

Scale Dropped Without Further Revisions or Trials

Due to the overwhelming, negative reviewer reaction, the scale was dropped before even being tried out on students.

Initial Versions



Summary of Instructions: The teacher put on the chalkboard: □ + △ = 10. You can see it in the upper left-hand corner of your page. She asked the students to think of pairs of numbers to put in the □ and the △ so that the equation would be true. She kept track of their answers in a table which she made right below the equation.

Each of the pairs worked. Let's check the first pairs, 6 and 4. If we put the 6 in the □ and the 4 in the △, we get 6 + 4 = = 10. Is that true? Yes, so 6 and 4 is a good pair for the table. Now take the next pair, 5 and 5. (Wait for students to think, then check it.) If you check the other pairs you'll see that they are good pairs also. The dots show that many, many different pairs will work. The next equation, beside the last one is □ x △ = 24. (Check the first two pairs with the students.)

Now at the bottom of the page are two more equations each with a table below it. For each equation, you think up as many good pairs for it as you can and record them in the table.

Reactions from Writers, Panelists and Coordinators

Two panelists thought the scale should be dropped. One thought the item biased toward CSMP students. The other thought the exercise not mathematically justifiable. A third panelist thought the scale acceptable only if the universe of acceptable answers was specified.

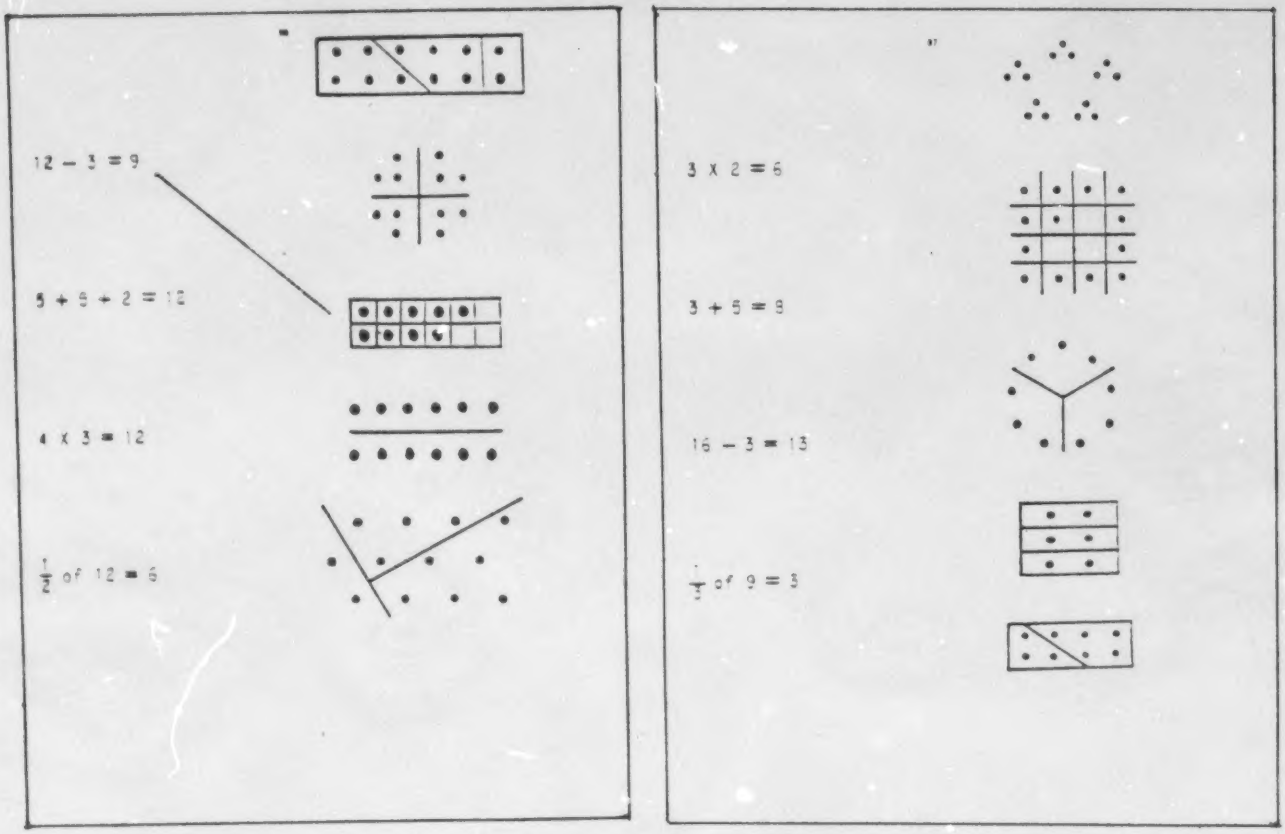
Decision to Drop

The scale was dropped without further development mainly due to the adverse reviewer reaction, particularly that of possible bias.

DOT PICTURES

Initial Version:

Student Pages:



Summary of Instructions: On this (the left hand) page there are four number sentences. The first one is 12-3=9, and a line has been drawn from that number sentence to one of the dot pictures on the other side of the page. That is because that dot picture is the best picture on this page for showing that 12-3=9. When I tell you to start, you have to do the same for the next three number sentences. Now after you have done that, there will be one picture left over. You are to make up a number sentence that goes with the picture that's left over, the one that didn't match with any number sentence. Just write your number sentence here on the blank at the bottom of the page. When you have finished this page just go on and do the next page.

Reactions from Writers, Panelists and Coordinators

Four reviewers found the scale unacceptable. One panelist found the items to be both unimportant and biased in favor of CSMP students. Another panelist found the scale technically hard to handle: correlated errors with resultant problems in scoring. Two writers simply didn't like this type of item. A third panelist found the scale acceptable but thought it should be revised considerably.

Decision to Drop

Due to the negative reviewer reactions (particularly the one having to do with bias), the scale was rejected at this time.

Scales Which Were Revised and Pilot Tested But Not Accepted

Iterative Frogs

Number Line Placing

Number Sentences About 12

Word Problems, Order Relations

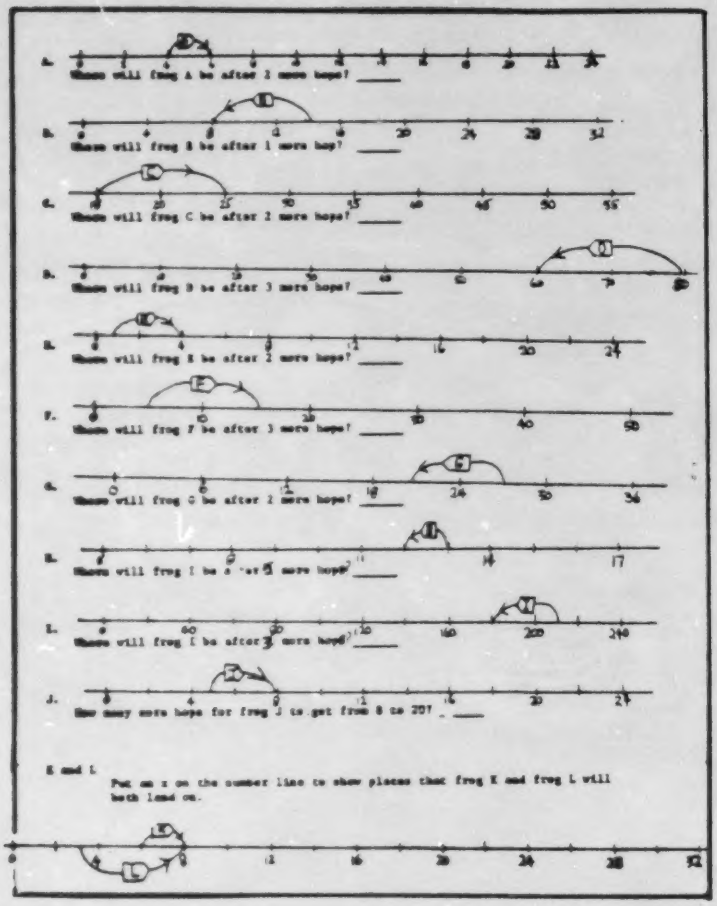
Word Problems, Extraneous Data

Rules and Examples

Oral Number Facts

Coins

Initial Version



Summary of Instructions: The "D" indicates a mechanical frog who always hops the same direction and distance each time: different letters, different frogs. First two examples worked out jointly and explained. Frog starts either on a mark or 1/2 way between two marks. Be careful of differing number lines.

Suggestions for improvement were (a) drop the items K and L because of the ambiguous wording and because of the increase in level of difficulty over item J, (b) include negative numbers on the number line (c) instruct the frog to go beyond the line (d) one writer thought the number line should be left off completely (thus bringing the bias criticism even more into play).

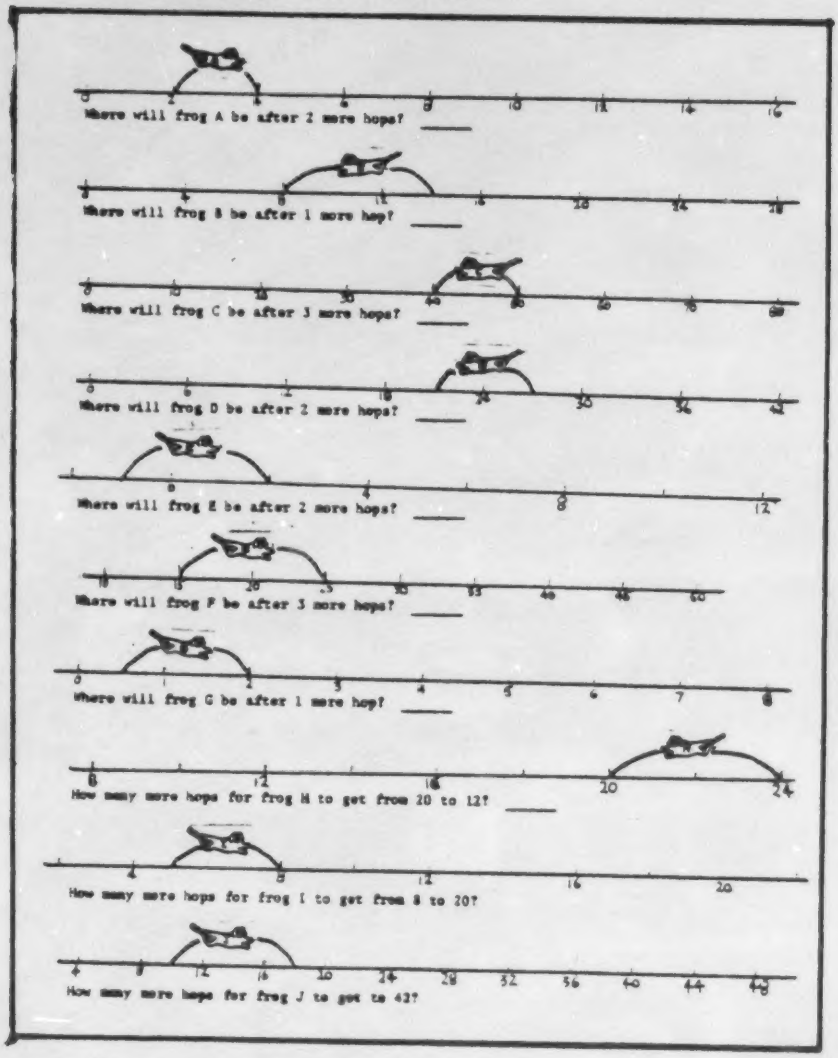
Reactions from Writers, Panelists and Coordinators

Two panelists and one writer found the scale to be unacceptable. One of the panelists thought it too biased in favo of the students studying CSMP with its arrow pictures. A panelist who otherwise found it acceptable, thought it very similar to the Number Line Labeling.

72

Revised Version

Items K and L were dropped. The E Frog was started at -1. The F frog was allowed to be off the illustrated line. The G frog started at 1/2 and was to land on 3½. The number of items of type J was increased from one to three.


Where will frog A be after 2 more hops?
Where will frog B be after 1 more hop?
Where will frog C be after 3 more hops?
Where will frog D be after 2 more hops?
Where will frog E be after 2 more hops?
Where will frog F be after 3 more hops?
Where will frog G be after 1 more hop?
How many more hops for frog H to get from 20 to 12?
How many more hops for frog I to get from 8 to 20?
How many more hops for frog J to get to 42?

Summary of Instructions: Same as previous except a warning was given about the last three items being different.

Trials with Students

After a trial with one pilot class it was observed that many students drew in hops from one mark to the next regardless of the indicated hop distance. In an attempt to discourage that behavior, the instructions were augumented to include a reminder that different frogs jump different amounts. In addition, further help was given in explaining the second example more thoroughly.

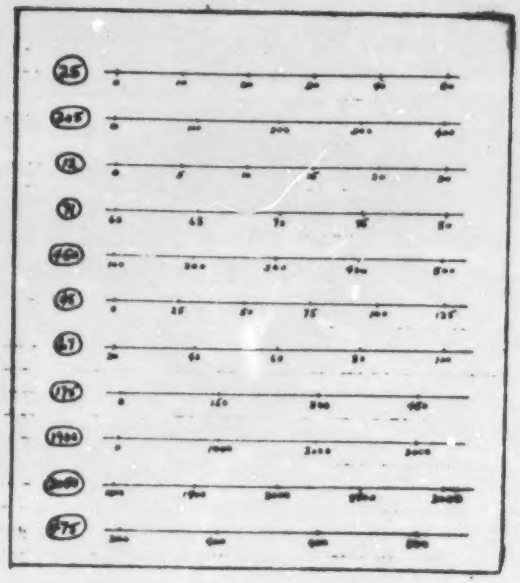
After trials with two more classes it was obvious that the task was very difficult for third graders to grasp. The average score was 1.9 out of a possible 8.0: only 1 of 65 students got a score of 6 or more items correct out of 8.

Decision to Drop the Scale

Due to the above results and the fact that possible bias had already been called into question, the scale was dropped.

NUMBER LINE PLACING

Initial Version:

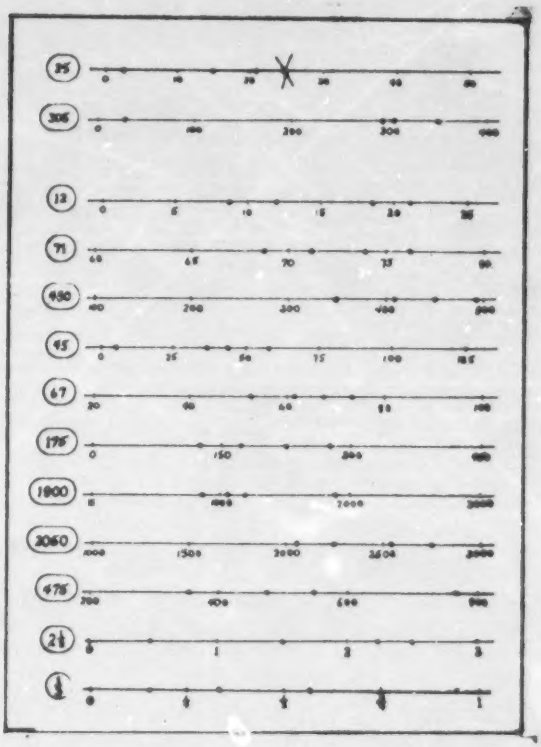


Summary of Instructions: For each number line, put an "x" to represent the circled number.

Reactions from Writers, Panelists and Coordinators

Almost universally favorable with a few minor suggested revisions: make scoring easier by making the answering procedure less open-ended, include items with negative numbers or fractions.

Revised Version:



Summary of Instructions: Same as initial version except four choices (dots) are given. The first (completed) item is explained. The second item is completed and explained.

Trials with Students

The revised version was tested with 3 classes of students. Some difficulty was encountered with describing unequivocably which was the correct dot for 305 (2nd practice item). The item analysis showed no particular weakness. Average score was 5.5 of 11.0.

Not selected

The scale was not selected for the final version because there were two scales involving very similar content, and the other one (Number Line Labeling) was easier to administer.

NUMBER SENTENCES ABOUT 12

Initial Version

Number Sentences

Mary's number sentences about 9.

$9 = 7 + 2$

$9 = 10 - 1$

$9 = 1 + 5 + 3$

My number sentences about 12.

12 = _____ 12 = _____

12 = _____ 12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

Summary of Instructions: Booklet page read to the students; each example explained. Can use extra space for more if you run out of lines.

Reactions from Writers, Panelists and Coordinators

The majority of reviewers had mild to serious objections to the scale. The mild objections involved poor wording in the instructions and the exclusive use of the +, -, and small whole numbers in the examples. There were three reviewers who found the scale unacceptable. A writer reported experience with classes having difficulty with this sort of activity. One panelist saw the task as "not a valid mathematical activity". Another thought the use of the phrase "true number sentences" unfair to students not familiar with it.

Revised Version #1

Accordingly, the scale was revised to meet most of the objections.

Number Sentences

Mary's number sentences about 9.

$9 = 7 + 2$ $9 = 3 \times 3$

$9 = 10 - 1$ $9 = 18 \div 2$

$9 = 1 + 5 + 3$ $9 = 4\frac{1}{2} \times 2$

My number sentences about 12.

12 = _____ 12 = _____

12 = _____ 12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

Summary of Instructions: Essentially the same as Initial Version.

Trials with Students

With the first pilot class it was found that very few students used other than + and - signs.

Revised Version #2

To highlight the possibility of the use of other than + and - signs, the order of the examples were switched and the instructions augmented.

Number Sentences

Mary's number sentences about 9.

$9 = 3 \times 3$ $9 = 10 - 1$

$9 = 7 + 2$ $9 = 1 + 5 + 3$

$9 = 18 \div 2$ $9 = 2 \times 4\frac{1}{2}$

My number sentences about 12.

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

12 = _____

Summary of Instructions: Instead of simply reading each sentence Mary wrote, the unique feature was pointed out (e.g., "in that example she is dividing").

The $4\frac{1}{2} \times 2$ example was changed to $2 \times 4\frac{1}{2}$. The two lines on the bottom right column were eliminated to facilitate the instructions about using the additional space if need be.

Trials with Students

Except for the fact that the previous pattern of responses continued with the second revision (and two more classes) the trials were quite successful. Most students were able to do the required task. The number of sentences produced varied widely.

Not Selected for Final Version

Since this scale and the one entitled "Equation Fluency" appeared to be very similar, only one was selected for MANS/77. Since the equation fluency scale received a slightly better reviewer reaction, it was selected.

Initial Versions

As noted previously each of five word problem scales grew out of a discussion about word problems in general. Before reactions to each scale could be received from the reviewers, a scale had to be developed to take advantage of the second round of pilot testing.

For the Reviewers

A. Peter is taller than John.
Dick is taller than John.
Who is taller; Peter or Dick?
Peter
Dick
Can't tell

B. Mary can run faster than Jane.
Tom is slower than Jane.
Who is fastest?
Mary
Tom
Jane
Can't tell

For the Student Booklet:

1. Peter is taller than John.

Dick is taller than John.

Who is tallest? Peter
Dick
John
Can't tell

2. Mary runs faster than Jane.

Tom is slower than Jane.

Who is fastest? Mary
Jane
Tom
Can't tell

3. Jack and Mike weigh 130 pounds together.

Jack and Pete weigh 100 pounds together.

Who weighs the most? Jack
Mike
Pete
Can't tell

4. Nobody in the class is taller than Maria.

Everybody in the class is taller than Fred.

Jane is in the class. Could she be

Taller than Maria?	YES	NO		Taller than Fred?	YES	NO
The same as Maria?	YES	NO		The same as Fred?	YES	NO
Shorter than Maria?	YES	NO		Shorter than Fred?	YES	NO

5. Jim and Bob weigh less than 100 pounds together.

Jim and Frank weigh exactly 100 pounds together.

Bob and Frank weigh more than 100 pounds together.

Who weigh the most? Jim
Bob
Frank
Can't tell

Reactions from the Reviewers and Pilot Trials

The reviewers' reactions were positive except for significant objections by two of the panelists. One thought the concept of a scale on order relations was good but did not lend itself well to good items. In addition he thought the scale was too heavily loaded toward classical logic.

The subjective reactions of the testers during the pilot trials were very negative. Item "4" continued to sound confusing in spite of minor revisions for the third trial class. Item "5" appeared to be very difficult for the students. The average score was a bit too high and the scale homogeneity was low.

Rejected from Final Version:

The subjective reactions of the testers during the pilot trials added to the strong negative reactions of two panel members lead to the rejection of this scale for the final version of MANS/77.

Note: Each problem to be read twice by the tester. Some added help was given on Item 1 (to get them started correctly) on item 4 because of unusual response format.

7.

WORD PROBLEMS: TYPE FOUR: EXTRANEOUS DATA

Initial Version

As noted previously each of five word problem scales grew out of a discussion about word problems in general. Before reactions to each scale could be received from the reviewers, a scale had to be developed to take advantage of the second round of pilot testing.

For the Reviewers

A. John had 31 marbles this morning. Mary had 19. During the day, John lost 10 marbles. How many did he have left at the end of the day? ___

Alternate A. John had 31 marbles this morning. Mary had 19. During the day, John lost 10 marbles. Circle the calculation that shows how many marbles John has at the end of the day.

19	31	31	31
-10	-19	+19	-10
9	12	50	21
	+10	-10	
	22	40	

B. Mary has 7 dolls to dress. Her dolls are each exactly 10 inches long. Each dress will cost 3 dollars. What will be the cost of dressing all 7 dolls? _____

Alternate B. Mary has seven dolls to dress. Her dolls are exactly ten inches long. Each dress will cost 3 dollars. Circle the calculation which Mary would use to figure the cost of dressing all seven dolls.

7	7	7	7
+3	x10	+10	x3
10	70	17	21

For the Student Booklets (2 Versions)

Version 1:

1. John had 13 marbles.
Mary had 11 marbles.
Then John lost 4 marbles in a game.
How many marbles did he have after the game?

A: 13 - 4 = 9

B: 13 - 11 = 2

C: 11 - 4 = 7

D: 13 + 11 = 24 , 24 - 4 = 20

2. Mary had 5 dolls to dress.
Her dolls are 6 inches long.
Each dress will cost 2 dollars.
How much will it cost to dress the dolls?

A: 5 + 2 = 7

B: 5 + 6 = 11

C: 5 x 6 = 30

D: 5 x 2 = 10

3. Peter went to the store with 4 friends.
Peter bought a candy bar for 10c.
He also bought a soda for 20c.
How much did Peter spend?

A: 4 x 10 = 40

B: 20 - 10 = 10

C: 20 + 10 = 30

D: 20 x 10 = 200

4. Jane gets 20c allowance each week.
She always saves half of her allowance.
So far she has saved 40c.
How much does she save each week?

A: ½ of 20 = 10

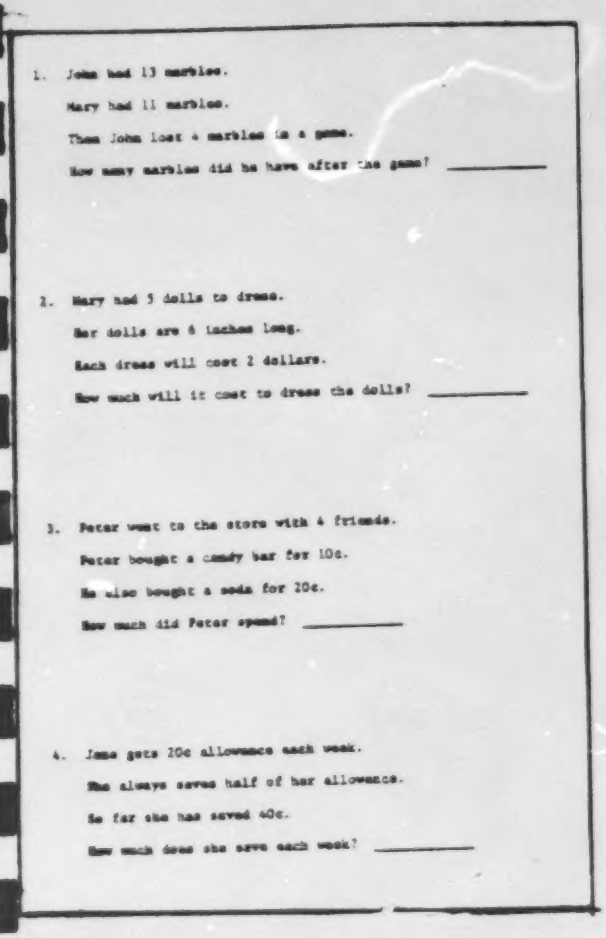
B: ½ of 40 = 20

C: 40 ÷ 20 = 60 , ½ of 60 = 30

D: 2 x 20 = 40

Note: Each version read twice by tester as student follows along. In the first version it is explained that each calculation has no errors.

Version 2:



1. John had 13 marbles.
   Mary had 11 marbles.
   Then John lost 4 marbles in a game.
   How many marbles did he have after the game? _____

2. Mary had 5 dolls to dress.
   Her dolls are 6 inches long.
   Each dress will cost 2 dollars.
   How much will it cost to dress the dolls? _____

3. Peter went to the store with 4 friends.
   Peter bought a candy bar for 10c.
   He also bought a soda for 20c.
   How much did Peter spend? _____

4. Jane gets 20c allowance each week.
   She always saves half of her allowance.
   So far she has saved 40c.
   How much does she save each week? _____

Reactions from Reviewers and Pilot Trials

Response from the reviewers was not generally positive. One panelist said he'd tried using problems with extraneous information and found that they produced little additional information over and above that obtained using word problems with no extraneous information. One writer could see no point in donfusing students with extraneous information. Several others had suggested changes: e.g., student could circle the extraneous data, language should be clarified, etc. All who expressed a preference, chose the first version.

Four classes were used in the pilot trials, two for each version. Students had problems dealing with the unannounced extraneous data: e.g. one student wanted to introduce additional information of his own creation. The item analysis showed a low measure of homogeneity for each version.
On each version the average score was 2.8 out of 4.0.

Not Selected for Final Version

Since the reviewers did not react favorably to this scale, since the trials did not give evidence to refute theoretical claims against it, and since there were other word problem scales being used which were as good or better, this scale was not selected for the MANS

RULES (AND EXAMPLES)

Initial Version

Tester Page

"I am going to read several sentences to you. After each one, you decide if it's true or false. If you think it's true, put a 'T' in the blank beside the letter. If you think it's false, put an 'F' in the blank.

Okay here is the first sentence, sentence 'A', (each sentence is read twice)

A. Add zero to a number and it will not change the number.

B. Multiply a number by zero and it will not change the number.

C. If you have two numbers, when you subtract the first from the second, the answer is always the same as when you subtract the second from the first.

D. If you have two numbers, when you add the first to the second, the answer is always the same as when you add the second to the first.

E. One times any number always equals one.

F. If a counting number is multiplied by ten, the answer always ends in zero.

G. If you subtract zero from any number, the answer is always that same number.

H. Three and one-half is between three and four.

I. Zero is less than any other number.

J. Adding 99 to any number is always the same as adding 100 and subtracting one.

K. If you subtract six from five you get an answer less than zero.

L. Multiplying a number by five is the same as writing the number down five times and adding them all together.

M. An odd number added to an odd number is always an odd number.

N. An odd number times an even number is always an even number."

Student Page

A. _____	H. _____
B. _____	I. _____
C. _____	J. _____
D. _____	K. _____
E. _____	L. _____
F. _____	M. _____
G. _____	N. _____

Reactions from Writers, Panelists and Coordinators

Half of the reviewers did not like this scale. Two of them (a writer and a panelist) also gave a reason: "the scale is too verbal". Many suggested changes were pointed out: (a) Rule I should be dropped since the universe for any number for 3rd graders is not clear, (b) The word "between" in Rule 4 is meaningless without a number-line, (c) Rule L should read "five times a number..." (d)in Rule F, "counting number" is not a familar phrase (e) use Rules involving the concept of multiples and the distributive principle.

First Revised Version

Except for the last suggestion all the others were heeded by either dropping the item or changing the wording. On the last suggestion it was decided that, for third graders, those concepts were too unfamilar.

Tester Page

"Okay, put down your pencils. Now I am going to read several sentences to you. After each one, you decide if it's true or false. If you think it's true, circle the 'T'. If you think it's false, circle 'F'. After you have circled 'T' or 'F', try to think of an example that shows you are correct. If you think of an example, put it in the blank. Okay here is the first sentence, sentence 'A'. (each sentence is read twice followed by a 30 second pause).

A. Add zero to any number and the answer will be the same number.

B. Multiply any number by zero and the answer will be the same number.

C. If you have two different numbers, when you subtract the first from the second, the answer is always the same as when you subtract the second from the first.

D. If you have two numbers, when you add the first to the second, the answer is always the same as when you add the second to the first.

E. One times any number always equals one.

F. If any whole number is multiplied by ten, the answer always ends in zero.

G. If you subtract zero from any number, the answer is always zero.

H. Adding 99 to any number is always the same as adding 100 and subtracting one.

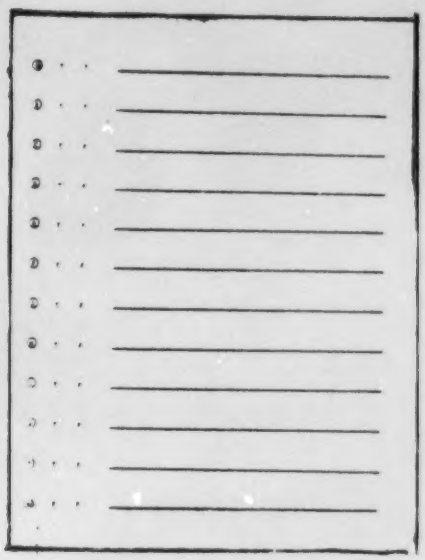
I. If you subtract a large number from a small number you get an answer less than zero.

J. Five times any number is the same as writing the number down five times and adding them all together.

K. An odd number added to an odd number is always an odd number.

L. An odd number times an even number is always an even number."

Student Page



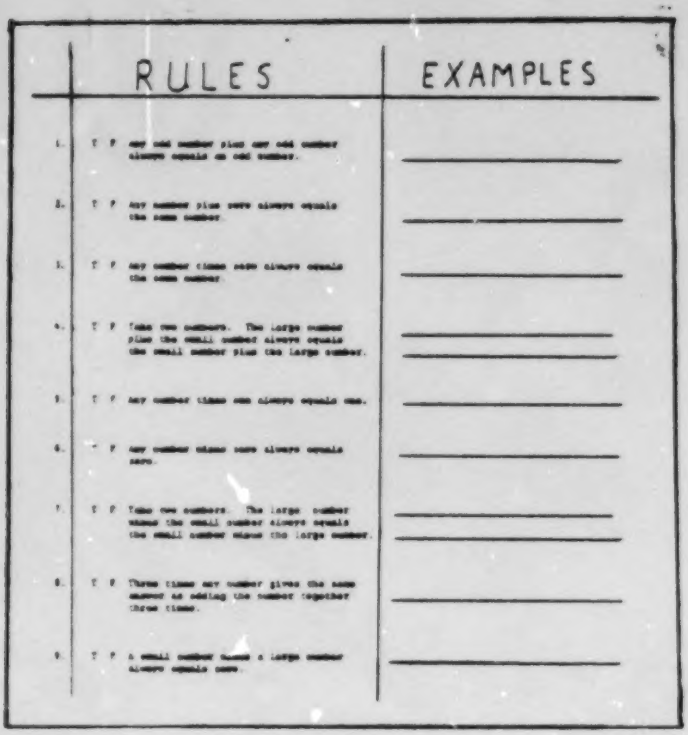
Trials with Students

After a trial with the pilot class it was obvious that the scale was too hard and too long. Hence for pilot classes two and three, item K was used as an example especially since it involved finding a true sentence to show that the "rule" was false. Items F, H, and L were dropped as being impossibly difficult: no student wrote any kind of an example for any of them.

The item analysis for the three classes showed that, for 5 of the 8 resulting items, the T and F score was close to 50%. In addition, the measure of scale homogeneity was extremely low. It seemed obvious that students were guessing: i.e., couldn't cope effectively with the task. Students still seemed to have difficulty with the example aspect, with true sentences showing rules to be false, and with rules which needed a two-sentence example (C & D).

Second Revised Version

In an attempt to help students cope with the task, the rules were written out in the student booklet; most of them were reworded. The word "EXAMPLES" was added above the lines and students were given 60 rather than 30 seconds to work on each one. To help alleviate the problem of a true sentence showing a rule to be false, the T and F were placed to the left of the rule; space for the example to the right. Two lines were provided when a two-sentence example was called for.



	RULES	EXAMPLES
1.	T F any odd number plus any odd number always equals an odd number.	_____
2.	T F any number plus zero always equals the same number.	_____
3.	T F any number times zero always equals the same number.	_____
4.	T F Take two numbers. The large number plus the small number always equals the small number plus the large number.	_____ _____
5.	T F any number times one always equals one.	_____
6.	T F any number minus zero always equals zero.	_____
7.	T F Take two numbers. The large number minus the small number always equals the small number minus the large number.	_____ _____
8.	T F Three times any number gives the same answer as adding the number together three times.	_____
9.	T F a small number minus a large number always equals zero.	_____

Summary of Instructions: Same as above: first item is worked out together, each rule is read twice.

Trials with Students

With two more pilot classes, evidence was still strong for guessing: 5 of 8 items close to 50% correct on T or F, scale homogeneity very low.

Decision to Drop Scale from MANS/77

Due to the overwhelmingly negative reaction from reviewers and the failure to salvage the scale with revisions, the scale was dropped.

ORAL NUMBER FACTS: "Type I, II and III"

Initial (and only) Version

Student Page

A. _____

B. _____

C. _____

D. _____

E. _____

F. _____

G. _____

H. _____

I. _____

J. _____

K. _____

Tester Page

Summary of Instructions: I'll read each question twice. Let's do question A together so you'll see how it goes.

Question A: How much is five plus seven?

B. Nine take away what number is three?

C. What number times seven equals twenty-one?

D. How much is half of fourteen?

E. Nine plus what number is thirteen?

F. What number minus seven equals four?

G. How much is six times three?

H. Half of what number is eight?

I. What number plus five equals eight?

J. How much is ten minus eight?

K. Three times what number equals fifteen?

Reactions from Writers, Panelists and Coordinators

Six of the eleven reviewers were in favor of dropping the scale. A writer thought it was merely a test of sharp listening. A panelist went further and called it "a test of one's ability to understand uninteresting questions expressed orally in an imperfect language." A coordinator thought the items not very novel.

Two suggestions for improvement were (1) Present the questions in written open sentence form (2) Add two items expressing questions involving subtraction in alternative language.

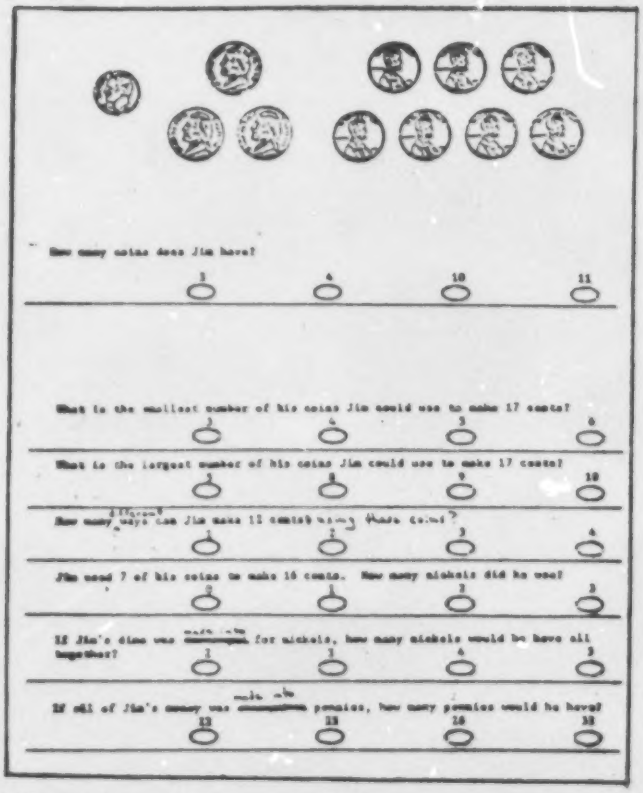
Trials with Students

With students the same initial version was used except that E was changed to "six plus what number is nine?" In the item analysis on trials with 3 classes, no clear differences appeared in the results of the questions of various types. The scale proved fairly easy for most students: average score 7.1 of 10.0.

Decision to Drop the Scale from MANS/77

The scale was dropped from the MANS/77 because it (a) received such a negative response from reviewers, (b) didn't appear to be accomplishing its purpose, and (c) would be prevented from being a discriminator by the ceiling effect.

Initial Version



Summary of Instructions: Tester tells about the coins Jim has and works through item #1 with the students. Students read and work items 2-7 on their own.

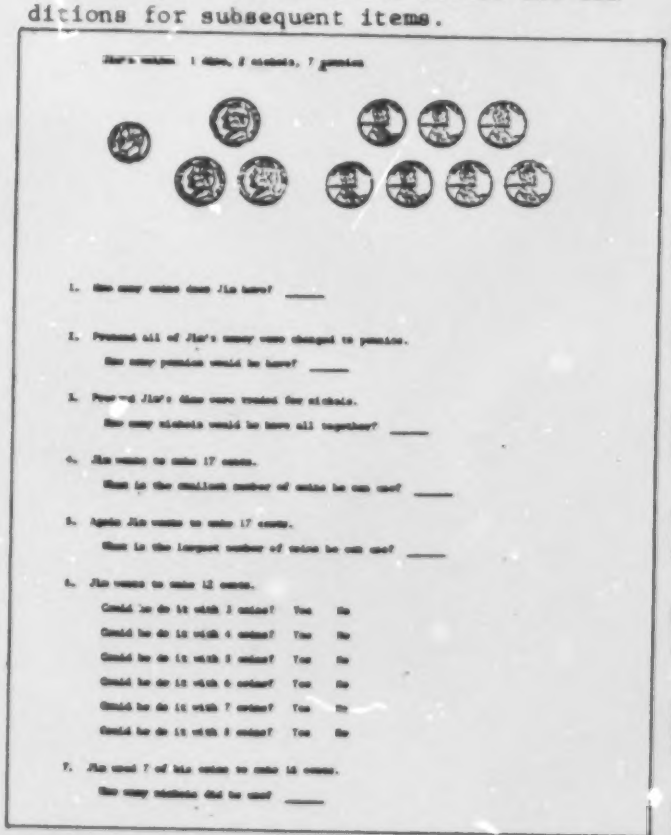
Reactions from Writers, Panelists and Coordinators

Even though the scale was acceptable to every reviewer, some expressed doubts about it and several suggested improvements. The doubts centered on whether or not students would be able to cope the items especially with the example being of so little help.

The suggested improvements were:
(1) either clarify "different" in item #3 or drop it as being too ambiguous.
(2)"replace by" or "exchanged for" rather than "made into" in items 5 and 6,
(3) place easiest items (5 and 6) at the beginning of the scale, (4) item 4 should say "...exactly 7...", (5) top of page should read "1 dime, 3 nickels, 7 pennies".

First Revised Version

In response to the worries about the scale being too hard, several changes were made: (a) the sentences were broken into shorter ones (b) new item 6 (old item 3) was put into a different format, (c) the multiple choice answer format was changed to completing the blanks, so as to avoid the by-chance floor effect of 25%. All the other suggested improvements were made or alleviated by other changes except for the "...exactly 7..." in new item 7: although technically correct it seemed to make the item harder. The word "Pretend" is needed in Items 2 and 3 so as to not confue the conditions for subsequent items.



Summary of Instructions: Same as above except that the different format of #6 is called to their attention.

Trials with Students

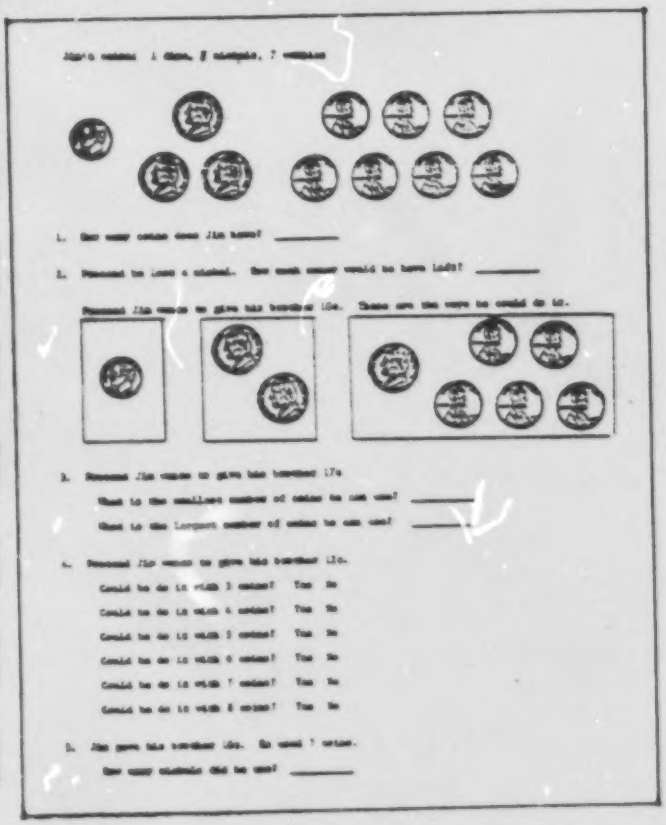
After a pilot trial with one class, it was clear that average third graders couldn't handle it. Hence for the trials with a second and third class, the tester read each item to the students and added two or three clarifying sentences after items 4-6.

8.1

The item analysis on these first three classes confirmed the observations. Since students had a 50% chance of being correct on the 6 parts of item 6, the average score of 5.1 out of 11.0 was deceptively high. The measure of scale homogeneity was very low. Students appeared to have trouble with the words "changed" and "traded" in items 2 and 3. Students did very poorly with items 4-7.

Second Revised Version

The problems with the words in Items 2 and 3 were alleviated by dropping them and adding a different "warm-up" type question of a similar nature (now #2). In order to help students with the last four items: (a) a "warm-up" exercise was inserted after Item 2, (b) Items 3 and 4 were treated more like a unit (new item #3).



Summary of Instructions: Essentially the same except for explaining the "warm-up" exercise. All items are read by the tester.

Trials with Students

After a trial with one class, it was observed that students were getting confused by the two sets of coin pictures. Hence for the second class, the three options for making up 10¢ were written in words.

The item analysis showed that many of the items were very difficult. Items 3 and 5 were still answered correctly by very few students. The measure of scale homogeneity was still too low.

Decision to Drop the Scale

Due to continued difficulties (above) and the fact that the scale was taking 10 to 12 minutes to administer, the scale was not used in MANS/77.

Appendix B

The MANS Scales

In this Appendix, a brief desciption of the scale, directions and time limits is given for each scale, together with a reduced version of the corresponding page(s) in the test booklets. The directions given here are merely outlines of the steps followed in the actual administration and hence are labelled "Abbreviated Directions". In fact the directions, sample items, explanations etc., were often fairly long; almost one-third of the total administration time was devoted to instructions and, in a sense, teaching of one kind or another. For this reason, and to insure that these directions were given accurately and consistently from one administration to another, the testers received extensive training for this task.

SCALE A1: Height and Weight Table (6 items)

ABSTRACT: Read and Interpret Data from a Table

ABBREVIATED DIRECTIONS:

Explanation of content of table
Oral questions followed by explanation of location of correct answer:
How tall was Sam last year?
How tall is Ann this year?
Sample questions P1, P2 (including the paraphrase "whose height went up 3 inches between last year and this year?")

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 3 minutes

WORKING TIME: Approximately 5 minutes

	LAST YEAR			THIS YEAR		
	Height (inches)	Weight (pounds)		Height (inches)	Weight (pounds)	
Joe	49	67		49	75	
Ann	41	63		47	71	
Peter	44	71		47	73	
Nancy	48	61		50	73	
Sam	42	75		46	79	
Cathy	46	77		47	82	

P1. Who was tallest last year? _____

P2. Who gained 3 inches in height? _____

1. Who weighs the most this year? _____
2. Who was shortest last year? _____
3. Who stayed the same height? _____
4. Who gained the most weight? _____
5. Who grew the most in height? _____
6. Who gained the least weight? _____

A2

SCALE A2: Estimation (25 items)

ABSTRACT: Estimate which of 5 intervals contains the answer to each of a series of computation problems. Three separate pages containing 8 addition, 8 subtraction and 7 multiplication problems respectively.

ABBREVIATED DIRECTIONS:

Caution not to figure out exact answers.
Referral to first practice example - emphasis on quick judgment of interval and placing "x".
Do second sample, with explanation of why the answer is less than 100 rather than greater than 100.
Start first page (A5) - work quickly, skip questions if desired.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 2½ minutes

WORKING TIME: Exactly 1½ minutes for each of pages A5, A7 and A9 - separately timed.

Practice

483 + 357 0 10 50 100 500 X 1000

48 + 49 0 10 50 100 500 1000

ADDITION

185 + 97 0 10 50 100 500 1000

24 + 24 0 10 50 100 500 1000

59 + 39 0 10 50 100 500 1000

479 + 86 0 10 50 100 500 1000

279 + 165 0 10 50 100 500 1000

19 + 29 0 10 50 100 500 1000

257 + 294 0 10 50 100 500 1000

19 + 19 + 19 0 10 50 100 500 1000

SUBTRACTION

100 - 93 0 10 50 100 500 1000

149 - 51 0 10 50 100 500 1000

137 - 125 0 10 50 100 500 1000

970 - 459 0 10 50 100 500 1000

260 - 211 0 10 50 100 500 1000

671 - 662 0 10 50 100 500 1000

700 - 205 0 10 50 100 500 1000

651 - 149 0 10 50 100 500 1000

MULTIPLICATION

5 X 109 0 10 50 100 500 1000

2 X 19 0 10 50 100 500 1000

40 X 10 0 10 50 100 500 1000

4 X 23 0 10 50 100 500 1000

11 X 50 0 10 50 100 500 1000

2 X 49 0 10 50 100 500 1000

4 X 29 0 10 50 100 500 1000

8.

SCALE A3: Functions (8 items)

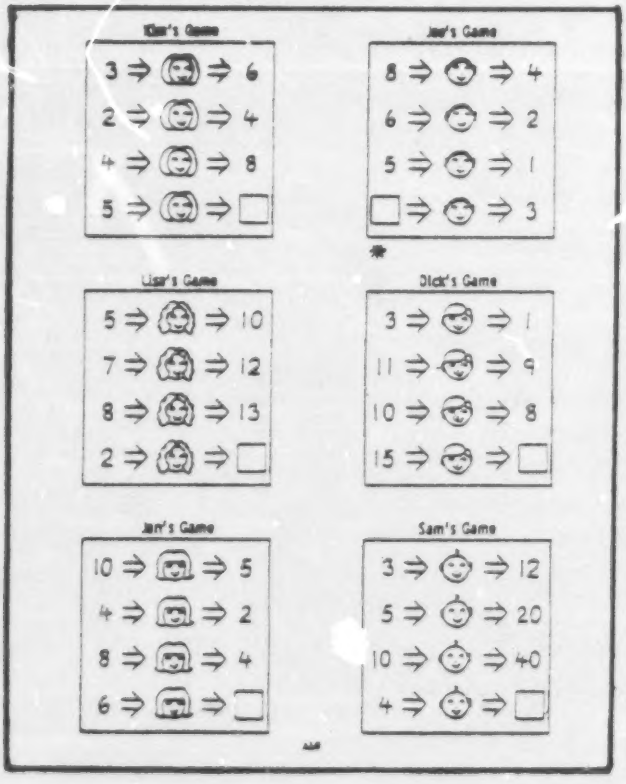
ABSTRACT: For each of several problems, determine from 3 pairs of numbers what the "secret rule" is which produces the second number from the first, and use it to find the missing number from the 4th pair.

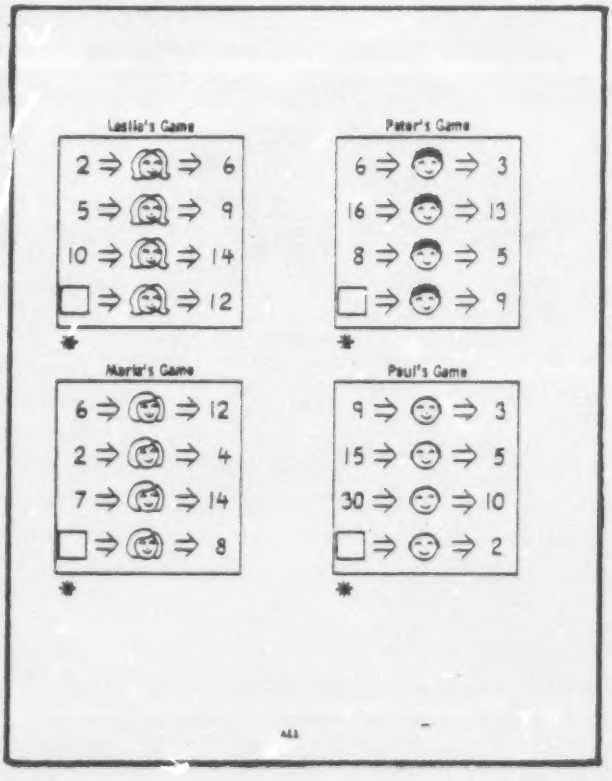
ABBREVIATED DIRECTIONS:

Referral to Kim's game
Possibility, from 3 ⇒ 6, of two rules (+3 or x2).
Determination from next two pairs (2 ⇒4,4 ⇒8) of rule "x2".
Calculation of missing number 5 x 2 = 10
Referral to Joe's game
Same steps as above with special emphasis on asterisk denoting that, for the last line, the number he started with is missing. Answer is 7 because 7-4 = 3, where secret rule is "-4".
Warning that questions on next page are of this type and that all rules are +, -, x, or ÷.

APPROXIMATE TIME REQUIRED FOR DIRECTIONS: 4½ minutes

WORKING TIME: Approximately 5½ minutes





SCALE A4: Two Stage Word Problems (5 items)

ABSTRACT: Word problems in which two different operations must be performed and where the numbers of the given data are relatively small.

ABBREVIATED DIRECTIONS:

Each question will be read twice and figuring can be done anywhere on the page.
Each question was then read twice to the students with an interval of approximately 1 minute between the second reading and the first reading of the next question.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 2 minutes (mostly reading time)

WORKING TIME: Approximately 5 minutes

1. Last year there were 25 rats on our island.
   Since then 5 rats have died and 3 were born.
   How many rats are on our island now? _____

2. On Saturday Amy and Susan made $13 selling lemonade.
   On Sunday they made $5.
   They put their money together and divided it evenly.
   How much did each girl get? _____

3. Our hens lay 9 eggs every day.
   Each day we eat 6 of them and give the others away.
   During the next 5 days how many eggs will we give away? _____

4. Jim has $10 in his bank now.
   Each week he will save $5.
   In how many weeks will he be able to buy a radio that costs $30? _____

5. There are 40 apples in our barrel now.
   We will eat 2 apples every day.
   How many apples will be left in our barrel after 5 days? _____

A13

9.

SCALE A5: Large Number Computations (12 items)

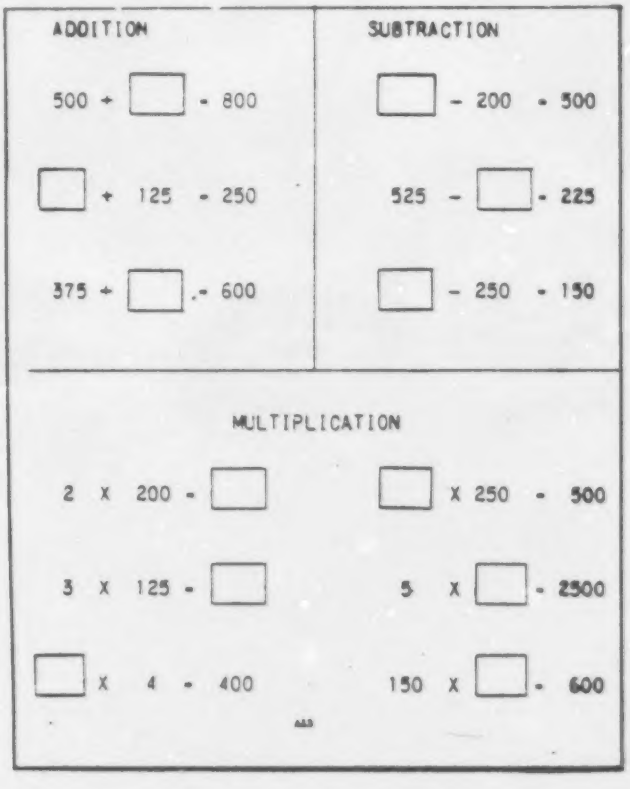
ABSTRACT: Solve computation problems given in an open sentence format, with placeholders in non-standard positions, and with numbers in the hundreds but relatively easy to work with (addition, subtraction and multiplication).

ABBREVIATED DIRECTIONS:

Find what goes in the box to make number sentence true.
Use empty spaces for figuring.
Go on if you get stuck.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: ¼ minute

WORKING TIME: Approximately 6 minutes



SCALE A6: Number Line Labelling (8 items)

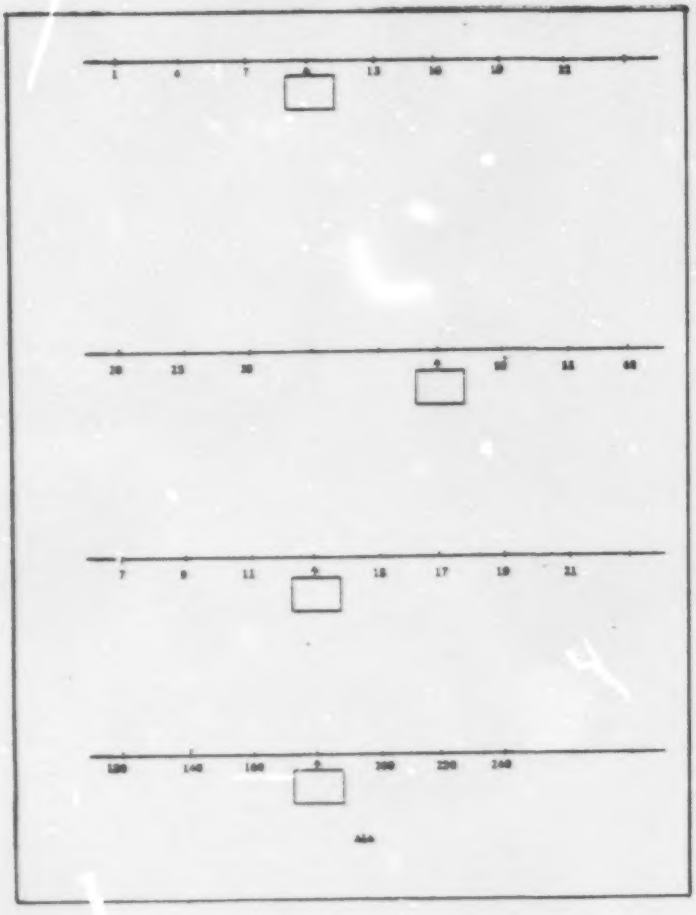
ABSTRACT: Label the indicated "mark" on several number lines, where marked intervals vary from item to item.

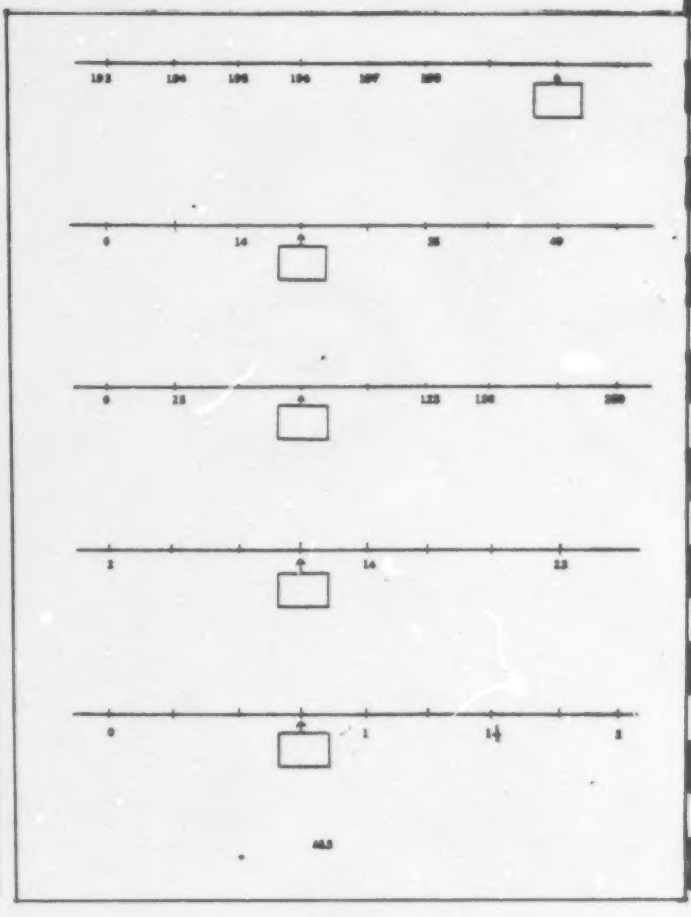
ABBREVIATED DIRECTIONS:

Review information given in sample (top of Page A14)
Numbers going up by 3 each time - complete sample.
Allowed to label any marks if desired.
Warning that intervals will vary from item to item.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 1½ minutes

WORKING TIME: Approximately 6 minutes





SCALE A7: Hints and Problems (5 items)

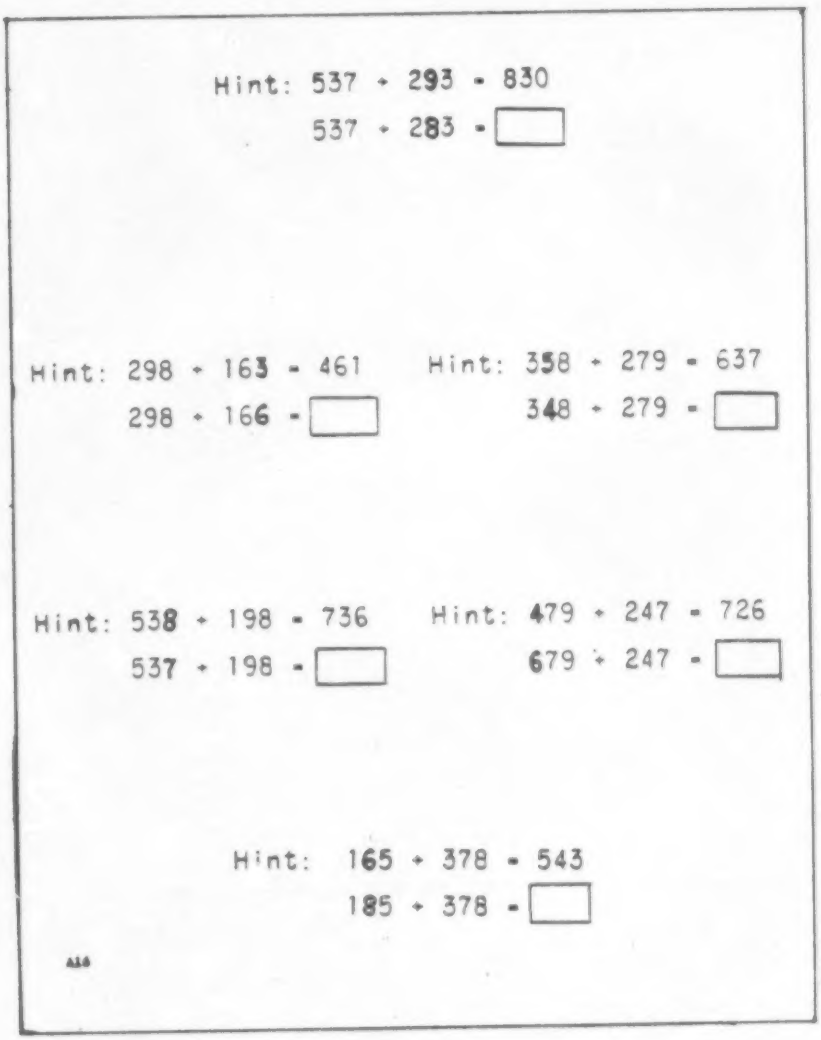
ABSTRACT: Complete a given addition problem by using the answer to another problem where one addend is the same as, and one is only slightly different from the given problem.

ABBREVIATED DIRECTIONS:

General idea of task
Review information in sample item (top of Page A16)
Difference between hint and problem, namely 293 versus 283, or 10 less
Answer will be 10 less, or 820
Warning about limited time - don't write out whole calculation but use hint.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 2½ minutes

WORKING TIME: Exactly 3 minutes



SCALE B1: 2 or 5 or 10 (10 items)

ABSTRACT: Estimate whether a given number is about 2 or 5 or 10 times as big as another given number

ABBREVIATED DIRECTIONS:

Explanation of task.
For sample item (top of Page B3), which of 2, 5 or 10 would best (though none is exactly right) fit the given sentence.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 1½ minutes

WORKING TIME: Exactly 2½ minutes (after 1¼ minutes students were told that half their time was up and they should be about half finished.)

2 or 5 or 10

60 is about ____ times as large as 31

51 is about ____ times as large as 5

16 is about ____ times as large as 7

65 is about ____ times as large as 12

200 is about ____ times as large as 21

100 is about ____ times as large as 19

98 is about ____ times as large as 50

45 is about ____ times as large as 8

602 is about ____ times as large as 298

503 is about ____ times as large as 49

59 is about ____ times as large as 31

SCALE B2: Composite Functions (9 items)

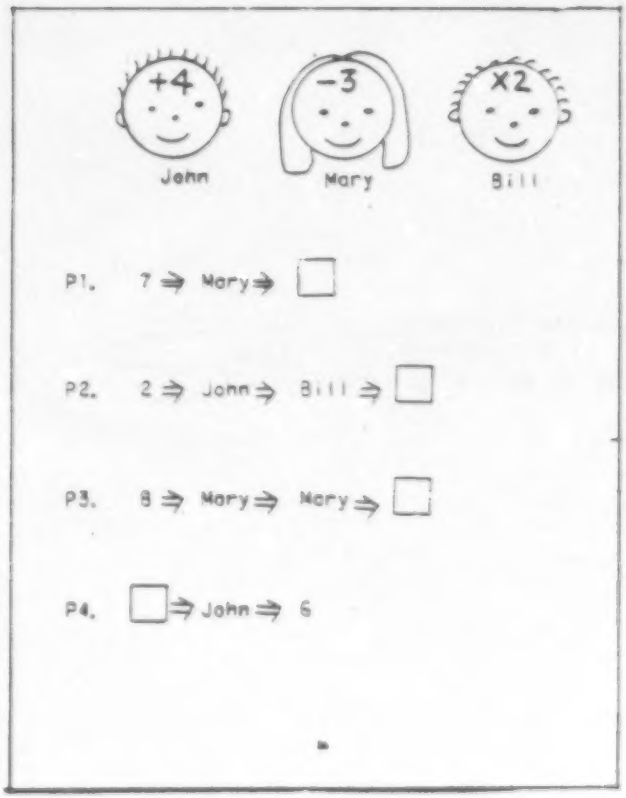
ABSTRACT: Starting with a given number, apply one or more operations in sequence and determine final result. Also, same process except final result is known and starting number is to be determined.

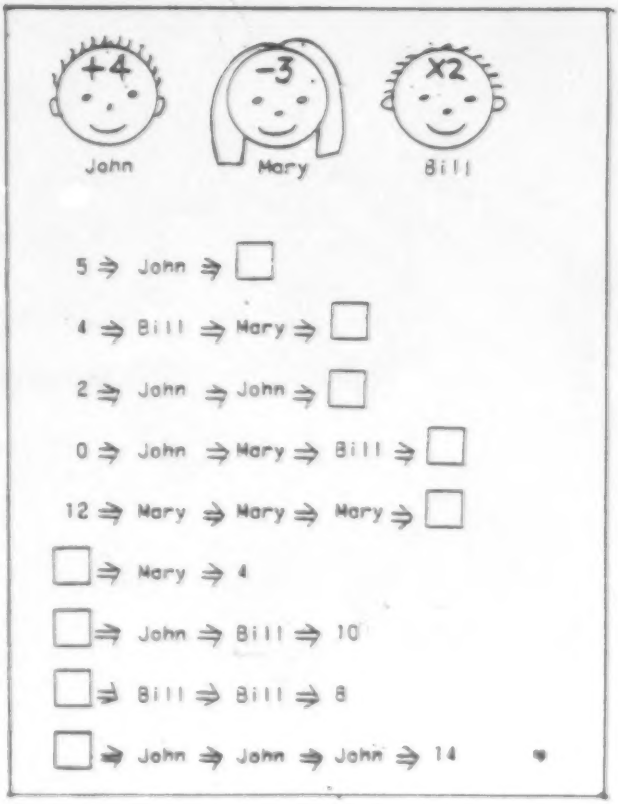
ABBREVIATED DIRECTIONS:

John, Mary and Bill have special calculators which only do one thing (as indicated)
Each sample item (P1-P4) done step by step
Special emphasis on last sample, which is similar to items at bottom of next page (B5), in that the starting number is to be determined.

APPROXIMATE TIME REQUIRED FOR DIRECTIONS: 5 minutes

WORKING TIME: Approximately 6 minutes




+4
John
-3
Mary
X2
Bill
5 ⇒ John ⇒ ☐
4 ⇒ Bill ⇒ Mary ⇒ ☐
2 ⇒ John ⇒ John ⇒ ☐
0 ⇒ John ⇒ Mary ⇒ Bill ⇒ ☐
12 ⇒ Mary ⇒ Mary ⇒ Mary ⇒ ☐
☐ ⇒ Mary ⇒ 4
☐ ⇒ John ⇒ Bill ⇒ 10
☐ ⇒ Bill ⇒ Bill ⇒ 8
☐ ⇒ John ⇒ John ⇒ John ⇒ 14

SCALE B3: Equation Fluency

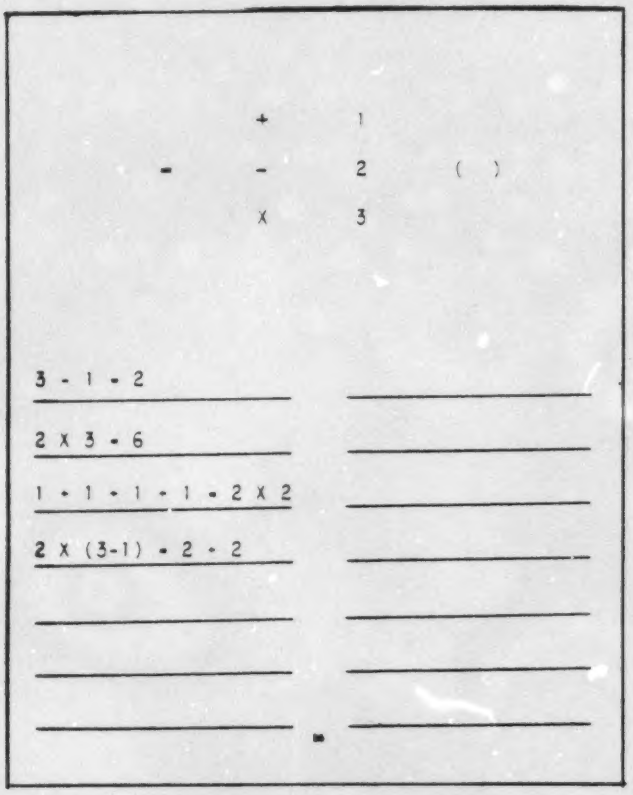
ABSTRACT: Given the symbols: = + - x 1 2 3 ( ), construct as many different equations as possible.

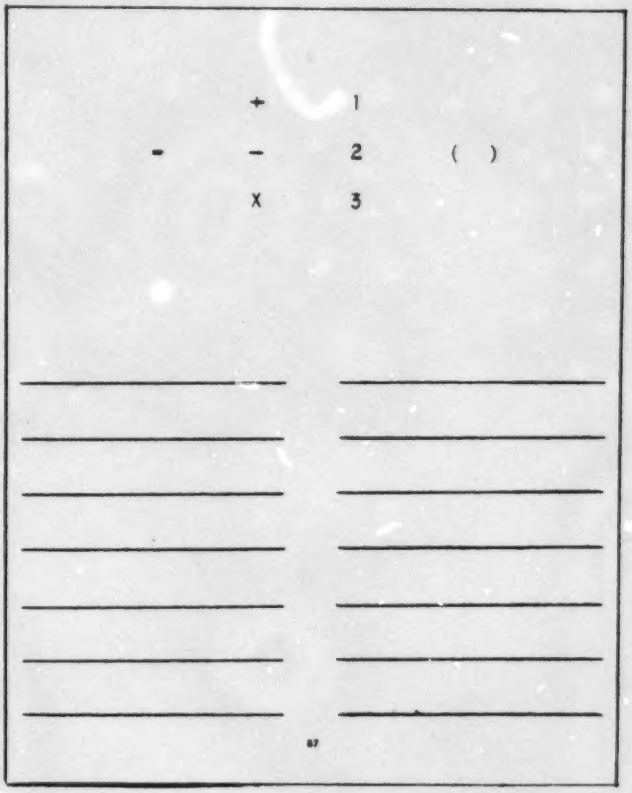
ABBREVIATED DIRECTIONS:

Read symbols at the top of the page.
May repeat symbols within a sentence.
Can't use symbols not given.
Can use parenthesis if desired
Review sample sentences, with special emphasis on non-allowable "2x3 = 6", because there is no "6" given at the top. (Students crossed this out and put in its place "1 x 3 = 3")
Do as many different ones as possible, may use next page.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 3 minutes

WORKING TIME: Exactly 5 minutes





SCALE B4: Circle the Larger (13 items)

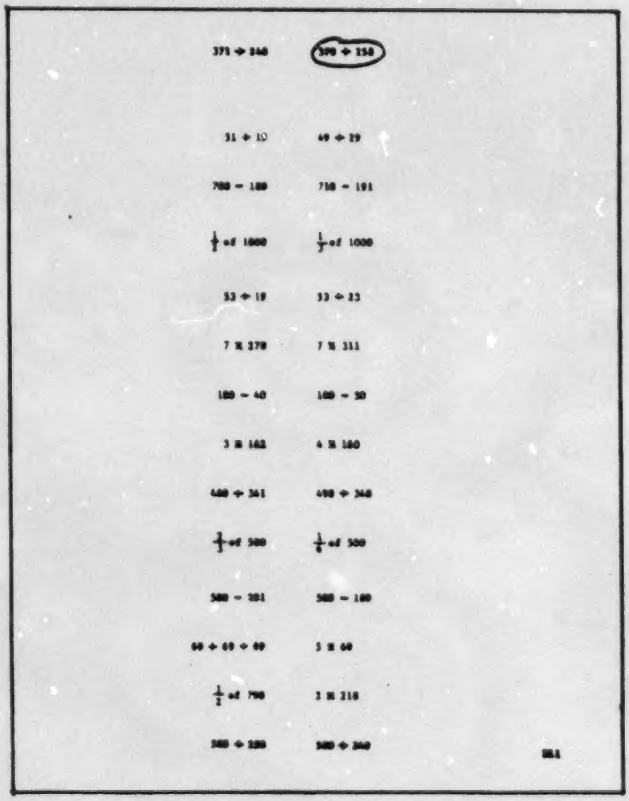
ABSTRACT: Given pairs of computation problems, determine which one has the larger answer.

ABBREVIATED DIRECTIONS:

Look at sample (Page B10)
371 is 1 more than 370, but 258 is 10 more than 248
So 370 + 258 is larger and is circled (a large demonstration card was used)
Do the same for each pair of problems (Page B11)
Don't try to calculate - not much time given.

APPROXIMATE TIME REQUIRED FOR DIRECTIONS: 1½ minutes

WORKING TIME: Exactly 3 minutes (after 1½ minutes students were told that half the time was up and they should be about half finished).



SCALE B5: Missing Digits (8 items)

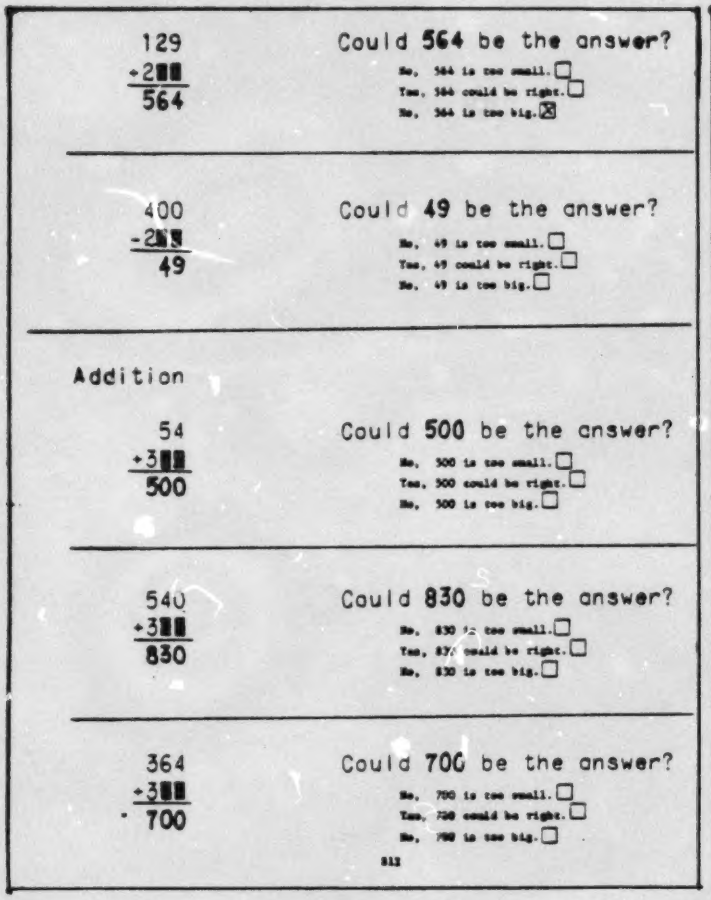
ABSTRACT: Given a computation problem with one or two digits of the problem part crossed out, determine whether or not the given answer could have been right (before the digits were crossed out).

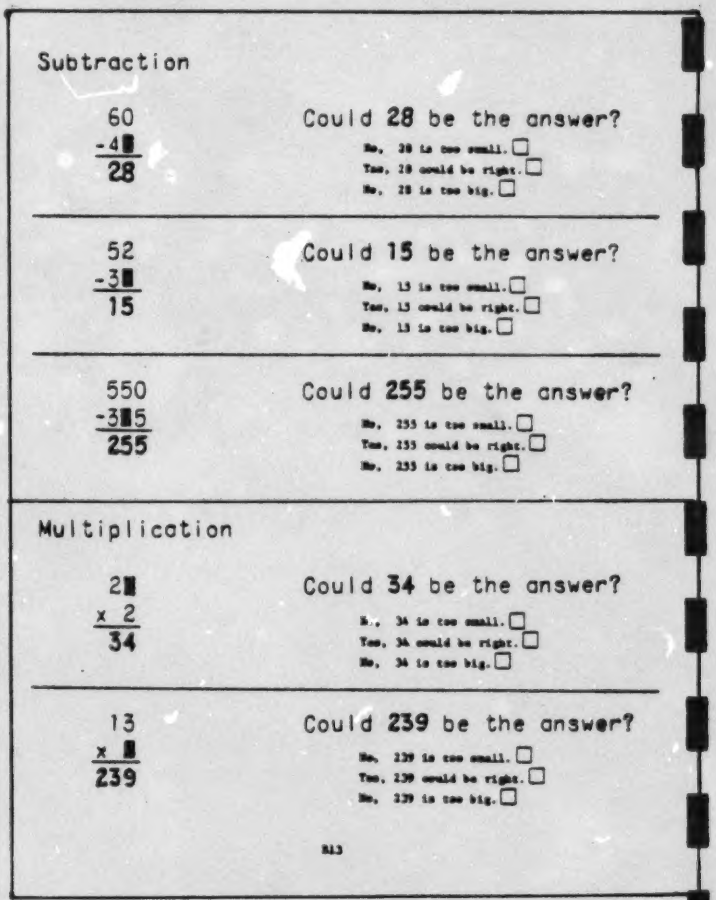
ABBREVIATED DIRECTIONS:

A student did the problems, then parts were covered up.
Look at first sample (top of Page B12).
Have to decide whether 564 could have been right.
Explanation of why it must be too big.
Second sample done, step by step.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 2½ minutes

WORKING TIME: 6 minutes





SCALE B6: Word Problems" with Rounding" (5 times)

ABSTRACT: Solve word problems involving division in which the given numbers do not divide evenly - (i.e., the answer, which must be an integer, can be obtained by rounding, up or down, the obtained quotient). The numbers of the given data were relatively small.

ABBREVIATED DIRECTIONS:

Each question will be read twice and figuring can be done anywhere on the page.
Each question was then read twice to the students with an interval of approximately 1 minute between the second reading and the first reading of the next question.

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 2 minutes (mostly reading time)

WORKING TIME: Approximately 5 minutes

1. An elevator can't hold more than 5 people.
   23 people want to ride to the top floor.
   How many times will the elevator have to go up? _____

2. Mr. Jones owns a car transporting company.
   His trucks can carry 3 cars each.
   He is loading 14 cars onto trucks.
   How many trucks will be need? _____

3. It takes 4 men to lift a piano.
   We have 14 men ready to work.
   How many pianos can they lift at the same time? _____

4. There are 18 hamsters in a pet store.
   The store owner is putting the hamsters into new cages.
   Each new cage holds no more than 4 hamsters.
   How many cages will he need? _____

5. A school wants to send as many volleyball teams as possible to a tournament.
   Each team must have 6 players.
   There are 21 volleyball players in the school.
   How many teams can they send? _____

813

SCALE B7: Fractions (8 items)

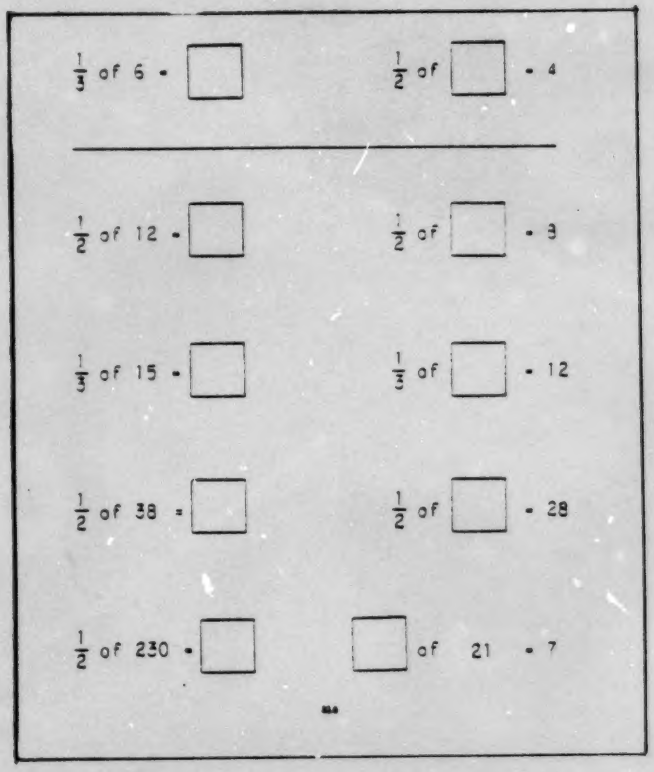
ABSTRACT: Solve problems of the form $\frac{1}{2}$ of y = ☐ or x of ☐ = y where x is 1/2 or 1/3.

ABBREVIATION DIRECTIONS:

Read sample (top left of Page B16)
Explain answers - analogy of dividing up 6 things among 3 people
Read sample (top right of Page B16)
Explain why "2" is not right answer.
Complete sample

APPROXIMATE TIME REQUIRED FOR INSTRUCTIONS: 1 and 3/4 minutes

WORKING TIME: Approximately 5½ minutes



10.

Appendix C

Statistical Data

This Appendix contains information on each MANS scales. There are usually two pages devoted to each subtest: the left-hand page provides numerical, especially item analysis data, and the right-hand page is a graph of the class mean scores.

Numerical Data

The content of each item is indicated very briefly (see Appendix B for actual student page) and beside each item is given:

a) the percent correct for all CSMP students and for all Non-CSMP students.

b) the r-biserial correlation between the item and the scale (a large biserial, i.e., one approaching 1.0 in size,* means the item is testing something very much the same as the rest of the items in the scale are testing; a small biserial, near 0, means the item is unrelated to the rest of the items), and

c) the common wrong responses in order of frequency, if given by more than 5% of the students (where more than 30% of the students gave the same wrong response, that percentage is given in parenthesis after the response).

Below the item data, the following information is given about the scale as a whole:

a) The correlation coefficient between the scale and each of the covariates.

b) The KR20 reliability coefficient. (A large KR20, near 1.0, means the items in the scales are testing a single underlying ability; a small KR20, near 0, means the items are rather unrelated to one another in what they are testing.)

c) The means by ability level. Each student was placed in one of five quintiles according to the percentile rank of his or her covariate score. Then the mean score for all students in the lowest (first) quintile was calcualated, similarly for students in the second, third, fourth and fifth quintiles. This was done separately for CSMP and Non-CSMP students and these quintile means are given in a table. Aside from the headings 1, 2..., 5 (for the quintile means) there is a heading "0" (for overall mean across all students) and "sd" (for overall standard deviation.)

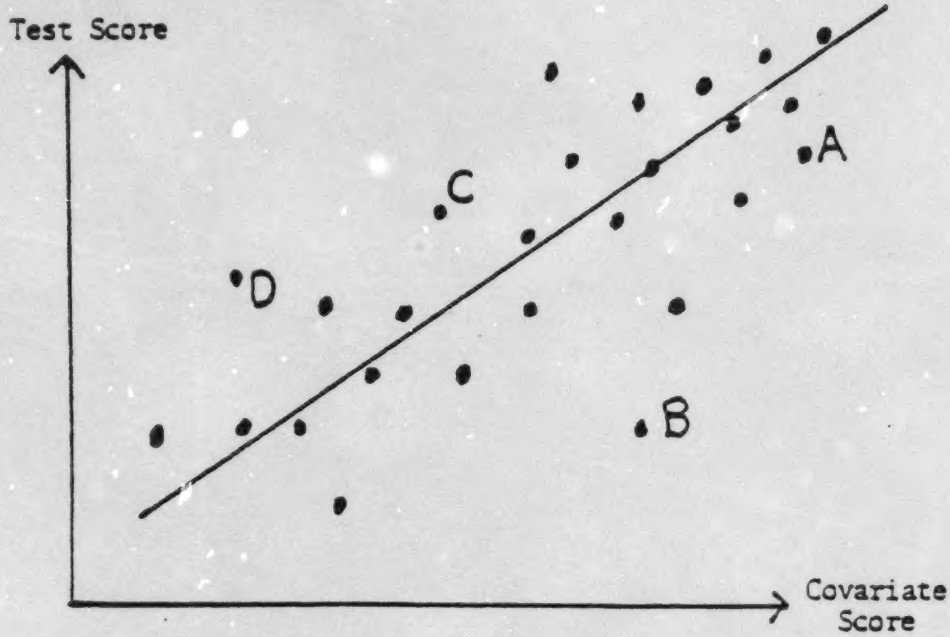
Below this scale data, a series of notes are given which point out or amplify some of the salient information.

The data was based on over 500 CSMP students and over 600 Non-CSMP students.

*The r-biserial may even exceed 1.0 with some data.

Graphs of Class Means

The graphs of class means are in the form shown below.

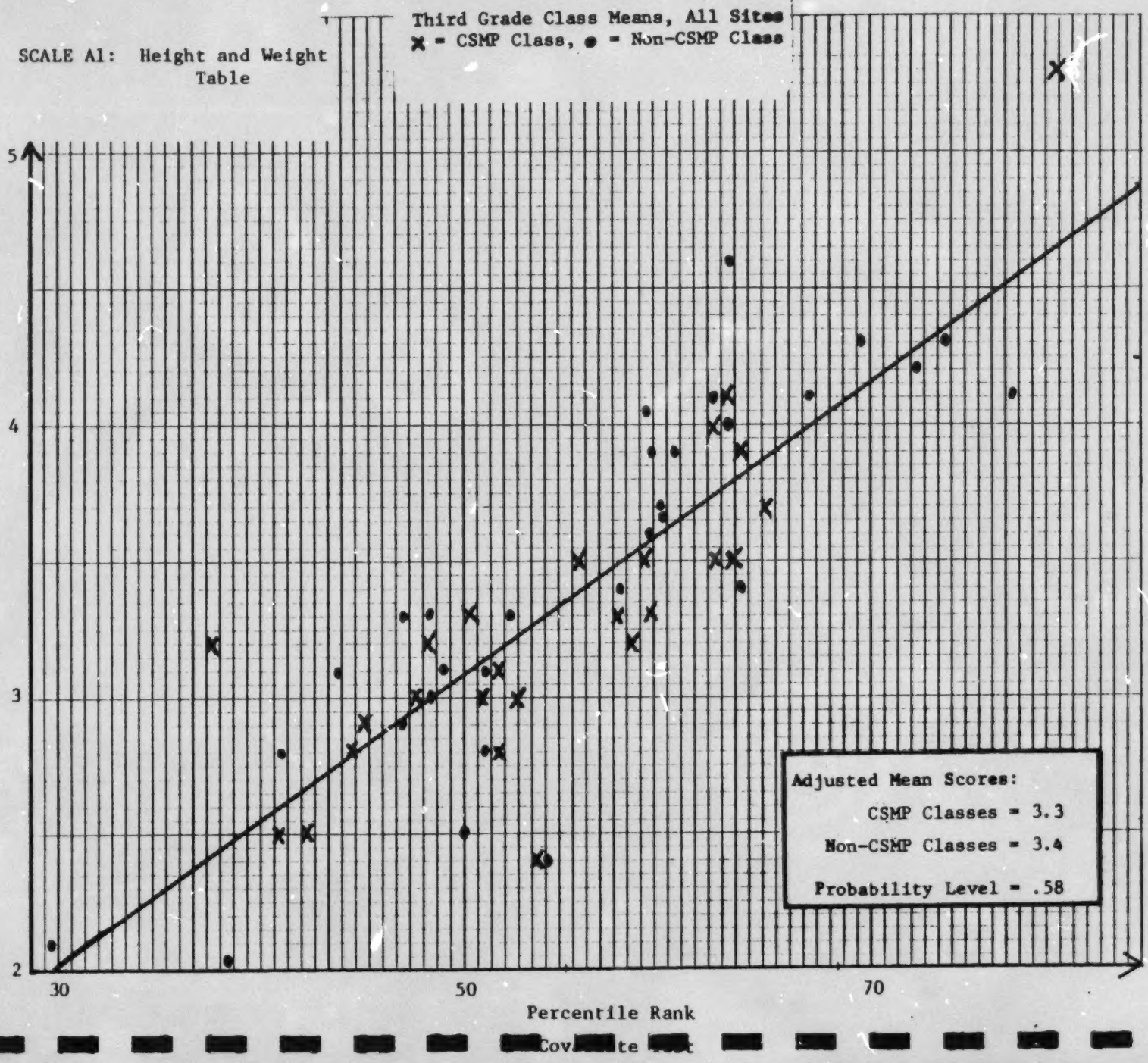


The vertical axis is for mean class test score on whatever test is being analyzed; the further towards the top of the graph - the higher the class score on that test. The horizontal axis is for covariate score, or mean class score on whatever covariate was used; the furtner to the right - the higher the general ability of the class. Based on the set of class means thus graphed, the regression line has been drawn. This line is the best linear prediction of mean class test score that can be made from knowing the ability level of the class. Note in the example that classes A and B fall well below the regression line, or are scoring well below what would be predicted for them from knowing the ability level of the class, while classes C and D fall well above the regression line. Note also that, although class A had a slightly higher mean score on the test than did class C, class C did much better given the relative ability scores of the two classes. When the class means generally fall close to the regression line, test scores are well predicted by the covariate; when they are more dispersed from this line, the covariate is a less effective predictor.

Throughout, an x is used to represent a CSMP class and a dot (·) represents a Non-CSMP class.

In an unused corner of each graph, a table gives the adjusted mean scores across CSMP classes and across Non-CSMP classes, together with the p-value obtained from the F-test. (The p-value is the probability that a difference in mean scores that large could occur by chance alone. Hence, a small p-value, especially below .05, indicates real differences between CSMP and Non-CSMP classes.)

As it turned out in the data analyzed here, the covariate was an effective predictor. The lowest percentage of variance accounted for by the covariate was 48% and the median for the 14 scales was 68%. The separate regression lines for CSMP and Non-CSMP are assumed to be parallel in this analysis. To test that assumption an F-test was run for each scale. The resulting p-value indicated that the assumption was justified; for none of the scales was the test significant at the .05 level.


Third Grade Class Means, All Sites
× = CSMP Class, ● = Non-CSMP Class
SCALE A1: Height and Weight Table
5
4
3
2
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 3.3
Non-CSMP Classes = 3.4
Probability Level = .58

Analysis of Scale A2

Estimation

Addition Items	Percent Correct		Biserial Correlation	Common Errors[1]
	CSMP	Non-CSMP		
General Format ADDITION 185 + 97 0 10 50 100 500 1000				
1. 185 + 97	76	69	.65	(500,1000), (50,100)
2. 24 + 24	77	74	.66	(50,100)
3. 59 + 39	72	69	.82	(100,500)
4. 479 + 86	52	50	.74	(100,500)
5. 279 + 165	53	52	.88	(500,1000)
6. 18 + 29	45	50	.93	(50,100)
7. 257 + 294	33	29	.74	(100,500)
8. 19 + 19 + 19	23	26	.64	(10,50)(100,500)

Note:

1. The percentages of students who did not complete items 4-7 were 15, 24, 32 and 40 respectively and these figures were higher for CSMP students than for Non-CSMP.

[1] "(500,1000)", for example, indicates that the interval between 500 and 1000 was a common (incorrect) response for the given calculation.

42 (continued)

Subtraction Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
[illegible]. 100 - 93	33	29	.71	(50,100)-32%,(10,50)(100,500)
10. 149 - 51	62	59	.67	(100,500),(10,50)
[illegible]. 137 - 125	40	36	.64	(50,100),(100,500),(0,10)
12. 970 - 459	26	38	.40	(100,500)-37%,(50,100),(10,50)
[illegible]. 260 - 211	19	22	.71	(50,100),(100,500),(500,1000)
14. 671 - 662	11	11	.74	(10,50),(50,100),(100,500),(50,1000)
[illegible]. 700 - 205	29	34	.49	(500,1000),(50,100)
[illegible]. 651 - 149	20	20	.43	(100,500),(50,100)

[illegible]te:

1. The percentages of students who did not complete items 13-15 were 17, 25, and 30 respectively with slightly higher percentages for CSMP students.

A2 (continued)

Multiplication Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
17. 5 x 109	58	50	.66	(100,500),(50,100)
18. 2 x 19	72	63	.64	(50,100),(100,500)
19. 40 x 10	48	41	.73	(50,100),(10,50),(500,1000)
20. 4 x 23	55	49	.68	(100,500)(10,50)
21. 11 x 50	33	26	.76	(100,500),(50,100),(10,50)
22. 2 x 49	52	49	.69	(100,500),(10,50)
23. 4 x 29	29	26	.60	(50,100)-35%,(10,50)

Correlations

KR 20 Reliability Homogeneity = .86

Kuhlmann Anderson = .56

ETS Cooperative Reading = .61

CTBS Computation = .43

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	7.0	8.8	9.7	11.0	14.3	10.2	5.4
Non-CSMP	6.9	8.6	9.4	10.4	12.6	9.6	5.3

Notes

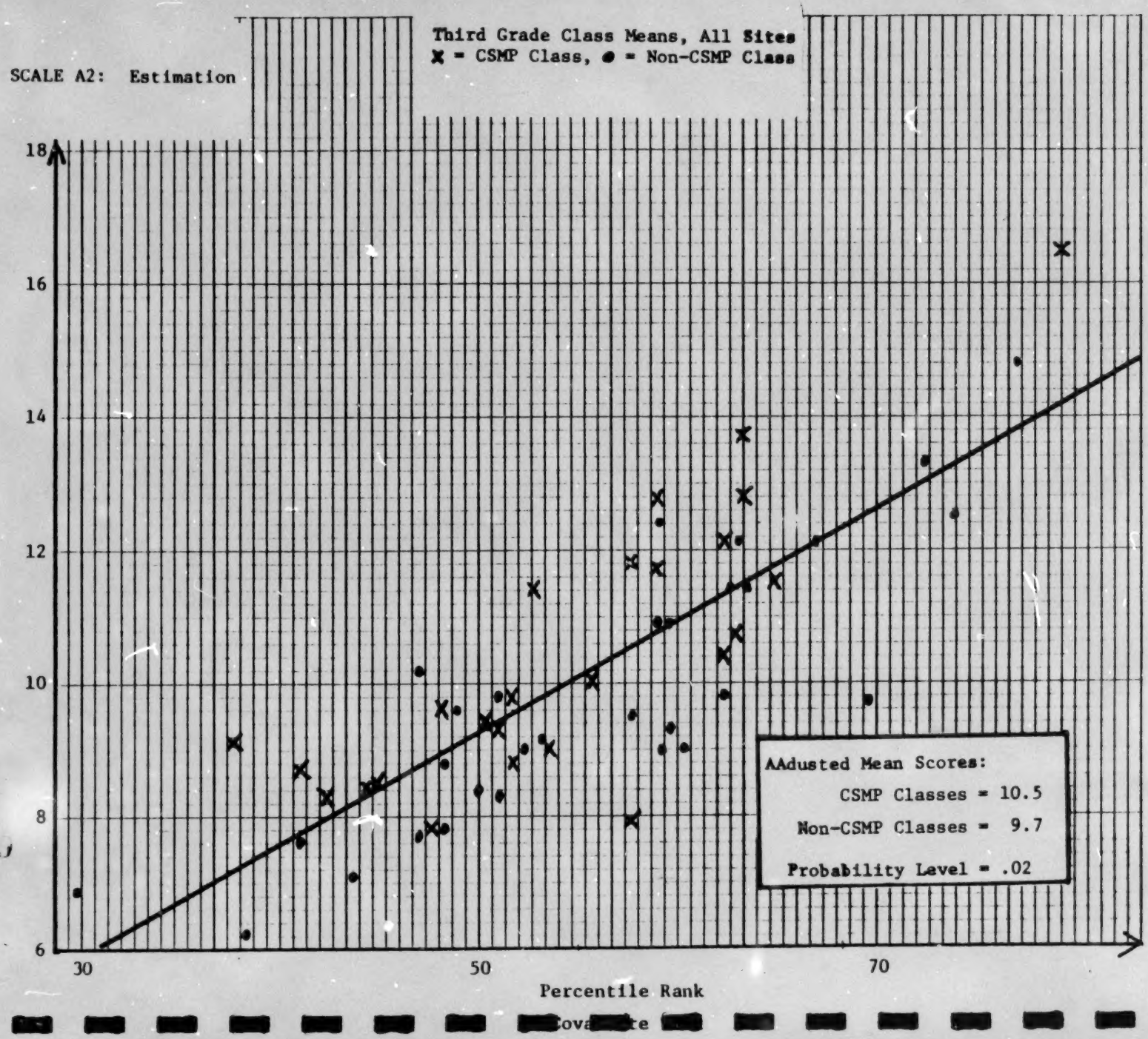
1. The percentages of students not completing items 21 and 22 were 11 and 15 respectively. Each scale was speeded, probably more than intended for the first two parts. Students generally completed successively more items with each scale; for example, about 85% of the students completed 4 addition items, about 83% completed 5 subtraction items, and 85% of the students completed 6 multiplication. Students were probably adjusting their rate of responding to the amount of time allowed.

2. The percentage correct listed for the later items in each part are misleading to some extent. For example, 33% of the CSMP students got item 7 correct, but 43% of the students omitted it. If one calculates the percent correct based on only those students who attempted the item, the percent correct would rise to 50%.

   When all the percentages are recalculated in this manner, the table below results. Wrong answers have been categorized into either "next best" or "other". (For example item 23, "4 x 29", (100,500) is correct, (50,100) is "next best" and all other intervals are "other").

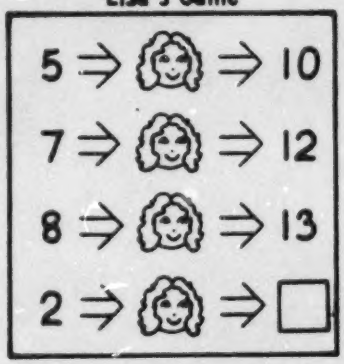
	Addition Items		Subtraction Items		Multiplication Items	
	CSMP Students	Non-CSMP Students	CSMP Students	Non-CSMP Students	CSMP Students	Non-CSMP Students
Mean Percent Correct	71	68	38	35	57	47
Mean Percent Next Best	21	23	27	27	25	26
Mean Percent Other	9	9	35	38	18	26

3. Looking at the table of means by ability level, one can see that the CSMP advantage increases rapidly as one progresses to the highest quintile. This is the one scale in which there appears to be an interaction between ability level and curriculum.


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE A2: Estimation
18
16
14
12
10
8
6
30
50
70
Percentile Rank
AAdusted Mean Scores:
CSMP Classes = 10.5
Non-CSMP Classes = 9.7
Probability Level = .02

Analysis of Scale A3

Functions

Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
1. General Format 				
1. (5,10) (7,12) (8,13) (2,□)	55	45	.87	4, 5
2. (3,1) (11,9) (10,8) (15,□)	65	52	.91	-2
3. (10,5) (4,2) (8,4) (6,□)	64	56	.73	-5
4. (3,12) (5,20) (10,40) (4,□)	38	37	.88	8, 13
5. (2,6) (5,9) (10,14) (□,12)	51	38	.92	6, 4
6. (6,3) (16,13)(8,5) (□,9)	38	34	.91	-3
7. (6,12, (2,4) (7,14)(□,8)	63	47	.79	2
8. (9,3)(15,5)(30,10) (□,2)	25	23	.79	4, 1

Correlations

KR 20 Reliability Homogeneity = .82

Kuhlmann Anderson = .63

ETS Cooperative Reading = .62

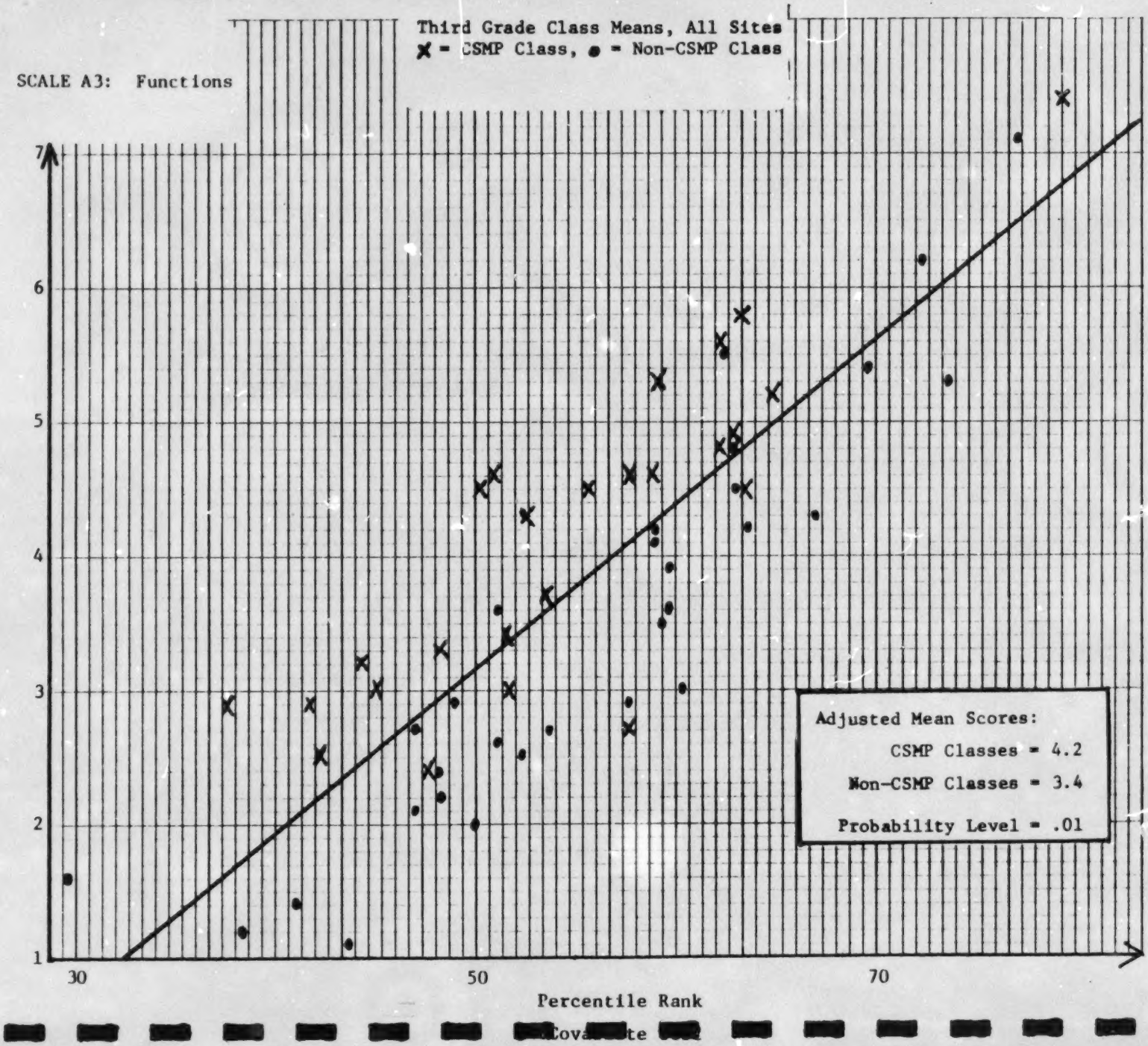
CTBS Computation = .52

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	2.0	3.1	4.0	4.9	6.1	4.0	2.5
Non-CSMP	1.2	2.3	3.0	4.0	5.8	3.3	2.6

Notes:

1. Across items, an average of about 5% of both the CSMP and Non-CSMP students gave a response using the wrong functions (for example, in #1 using the (5,10) to get the function 2x, instead of +5, thereby getting "4" in the box).
2. About 5% of the Non-CSMP students, and half as many CSMP students, tried to write the actual rule. (For example, in #1, "5" or "+5" was put in the box).
3. About 23% of the CSMP students, and 31% of the Non-CSMP students got no more than one item correct.
4. About 15% of the students did not complete Item 7, and 12% of the students did not complete Item 6


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE A3: Functions
7
6
5
4
3
2
1
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 4.2
Non-CSMP Classes = 3.4
Probability Level = .01

Analysis of Scale A4

Two Stage Word Problems

Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
1. Last year there were 25 rats on our island. Since then 5 rats have died and 3 were born How many rats are on our island now? ___	82	81	.84	
2. On Saturday Amy and Susan made $13 selling lemonade. On Sunday they made $5. They put their money together and divided it evenly. How much did each girl get?____	63	52	.86	18, 8
3. Our hens lay 9 eggs every day. Each day we eat 6 of them and give the others away. During the next 5 days how many eggs will we give away? _____	47	47	.88	3
4. Jim has $10 in his bank now. Each week he will save $5. In how many weeks will he be able to buy a radio that costs $30? ___	42	41	.78	6, 3
5. There are 40 apples in our barrel now. We will eat 2 apples every day. How many apples will be left in our barrel after 5 days?_____	57	47	.86	10

Correlations

KR 20 Reliability Homogeneity = .67

Kuhlmann Anderson = .49

ETS Cooperative Reading = .58

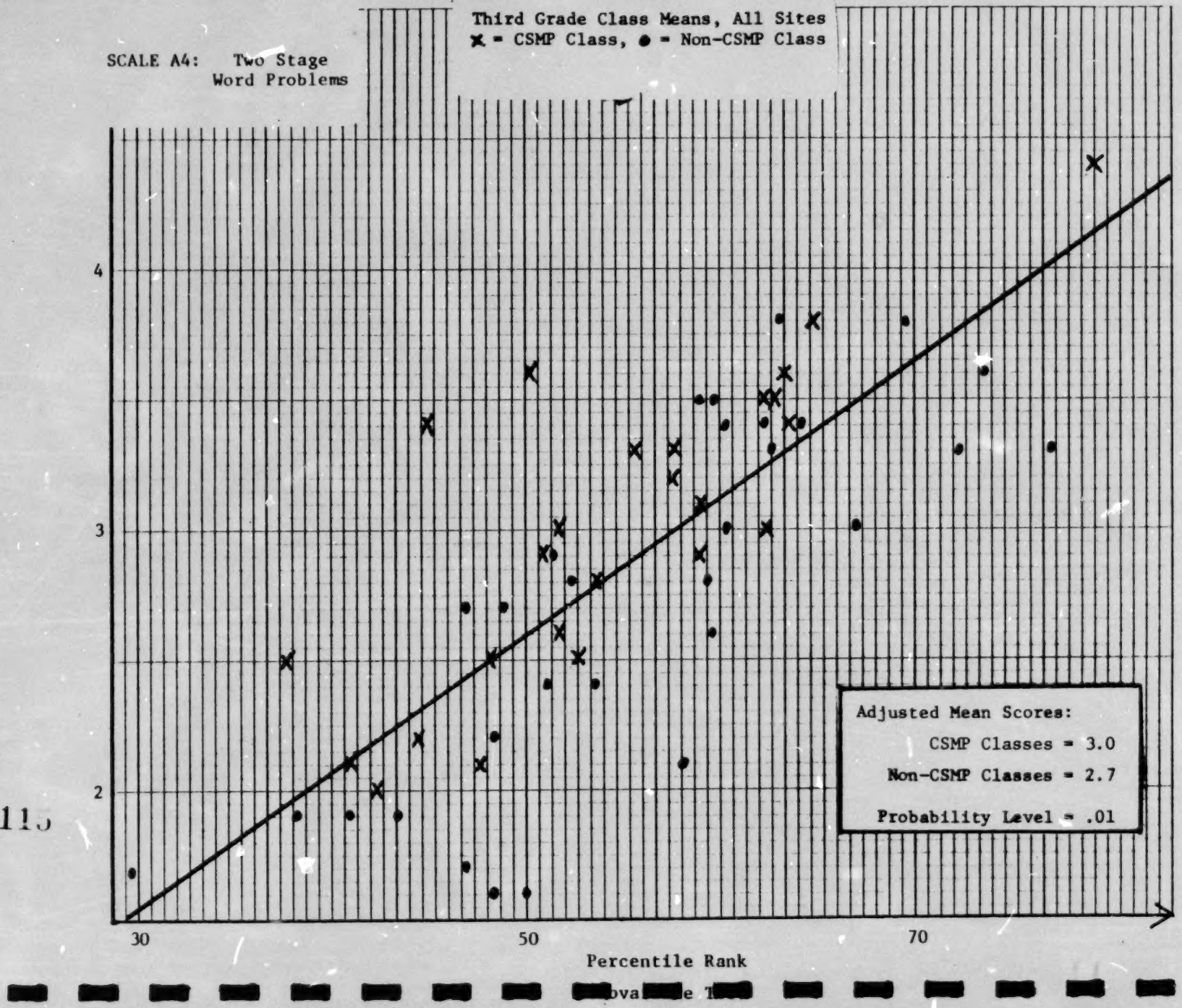
CTBS Computation = .50

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	1.9	2.5	2.9	3.4	4.0	2.9	1.5
Non-CSMP	1.6	2.2	2.7	3.0	3.8	2.7	1.6

Notes:

1. It can be seen that the small, but consistent, advantage enjoyed by the CSMP classes was due almost entirely to items 2 and 5.


Third Grade Class Means, All Sites
× = CSMP Class, ● = Non-CSMP Class
SCALE A4: Two Stage
Word Problems
4
3
2
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 3.0
Non-CSMP Classes = 2.7
Probability Level = .01

Analysis of Scale A5

Large Number Computation

Items	Percent Correct		Biserial Correlation*	Common Errors
	CSMP	Non-CSMP		
1. 500 + ☐ = 800	88	84	.98	
2. ☐ + 125 = 250	61	53	1.02	
3. 375 + ☐ = 600	25	24	.96	325
4. ☐ - 200 = 500	66	62	.96	300
5. 525 - ☐ = 225	50	43	.96	
6. ☐ - 250 = 150	08	10	.87	100
7. 2 x 200 = ☐	84	77	.81	
8. 3 x 125 = ☐	57	42	.90	75, 325
9. ☐ x 4 = 400	51	43	.85	
10. ☐ x 250 = 500	32	25	.96	250, 750
11. 5 x ☐ = 2500	20	18	.85	
12. 150 x ☐ = 600	19	14	.98	2

*Correlated with subpart, not total scale (where subparts are items 1-3, 4-6, 7-12)

Correlations

KR 20 Reliability Homogeneity = .80

Kuhlmann Anderson = .63

ETS Cooperative Reading = .71

CTBS Computation = .62

Means by Ability Level

	1	2	3	4	5	0	s[illegible]
CSMP	2.9	4.8	5.7	6.7	8.0	5.6	2.9
Non-CSMP	2.2	4.0	4.8	6.2	7.4	5.0	2[illegible]

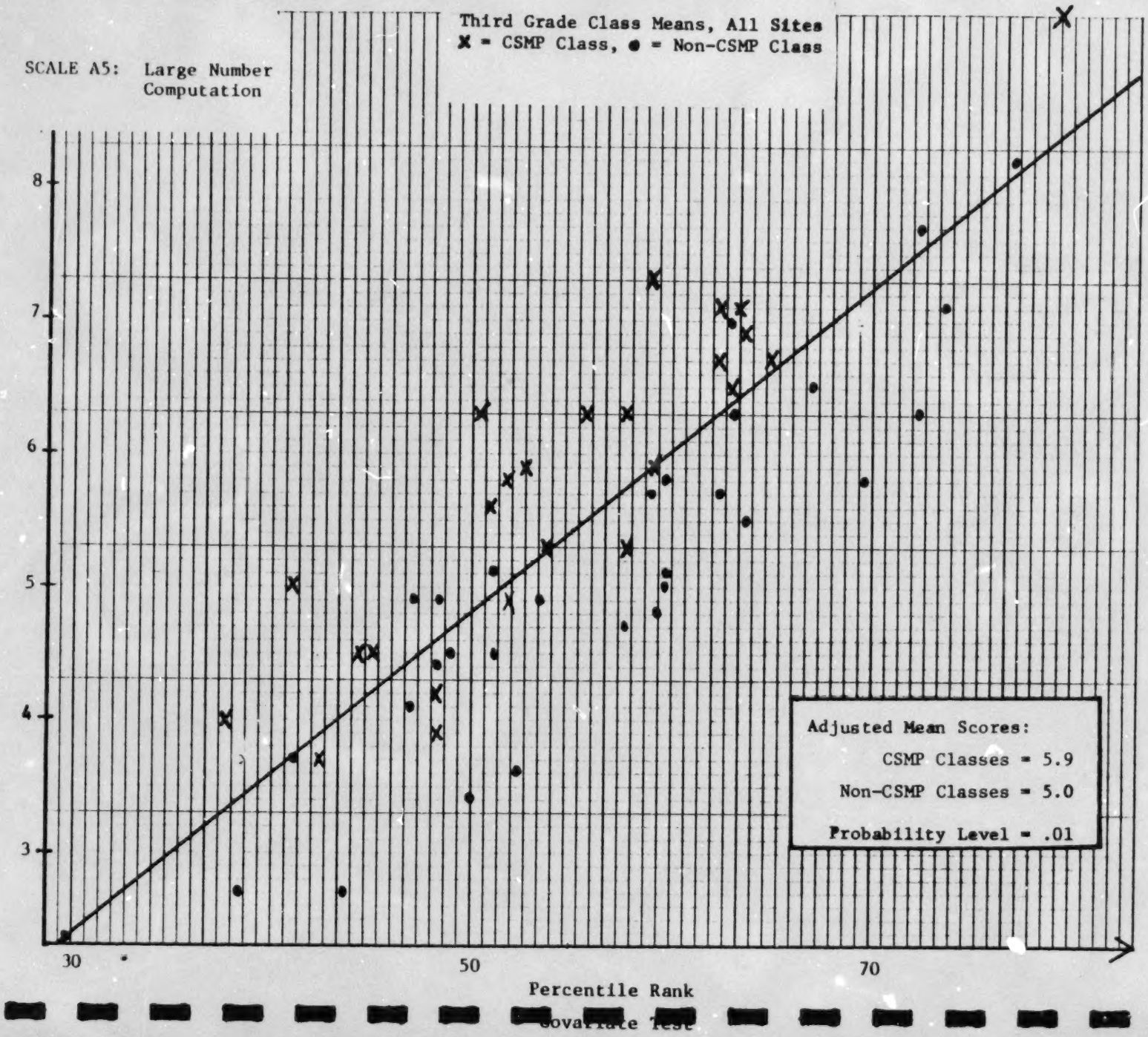
Notes

1. Most of the difference between CSMP and Non-CSMP was acccunted for by the multiplication items (7-12) where the mean scores were 2.6 versus 2.2.

2. About 19% of the students did not complete item 11, and 12% did not complete item 10. These figures were slightly higher for CSMP than for Non-CSMP students.

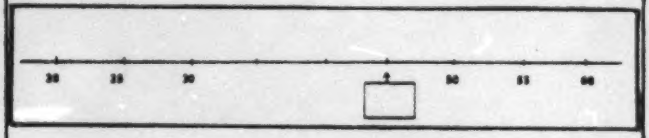
3. About 10% of the students got none of the three addition items (1-3) correct and 25% of the students got none of the subtraction items (4-6) correct. On the six multiplication items (7-12) 27% of the CSMP students, versus 41% for Non-CSMP, got <u>at most</u> one item correct.

4. A5 was "technically" one of the best scales: high reliability; the covariate was a very effective predictor; and the regression lines for CSMP and Non-CSMP classes were parallel (and can be easily visualized on the facing page).


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE A5: Large Number Computation
8
7
6
5
4
3
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 5.9
Non-CSMP Classes = 5.0
Probability Level = .01

Analysis of Scale A6

Number Line Labelling

Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
General Format 				
1. 20,25,30,__,__,☐,50,55,60	87	89	1.00	
2. 7,9,11,☐,15,17,19,21,__	87	87	.96	12 or 14
3. 120,140,160,☐,200,220,240,__,__	76	78	.91	170 or 190
4. 193,194,195,196,197,198,__,☐,__	62	61	.81	199
5. 0,__,14,☐,__,35,__,49,__	53	53	.90	28
6. 0,25,__,☐,__,125,150,__,200	46	41	.89	50 or 100
7. 2,__,__,☐,14,__,__,23	40	27	.70	10 or 12
8. 0,__,__,☐,1,__,1½,__,2	02	09	.58	1/4 or 1/3

Correlations

KR 20 Reliability Homogeneity = .75

Kuhlmann Anderson = .62

ETS Cooperative Reading = .60

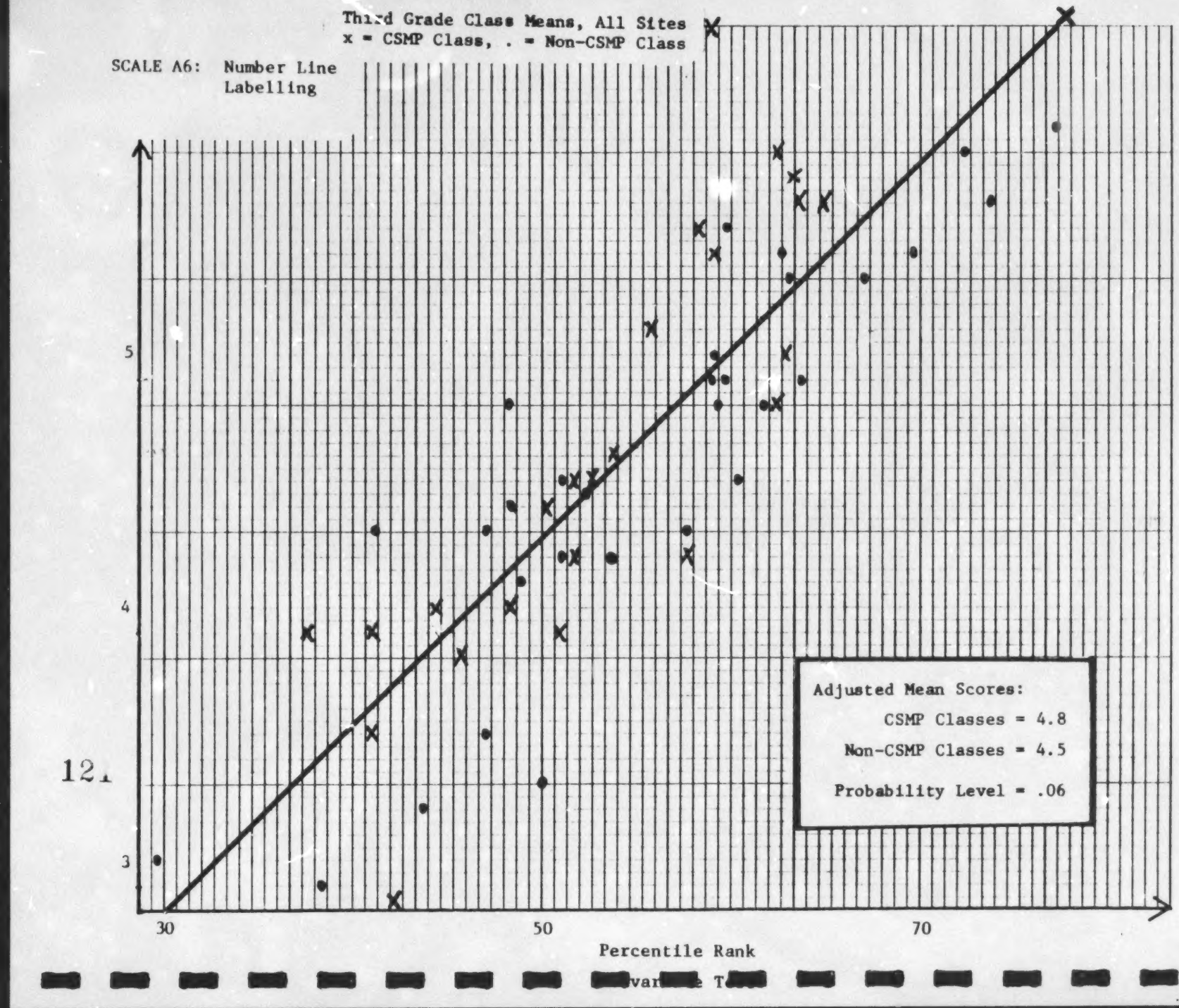
CTBS Computation = .58

Means in Ability Level

	1	2	3	4	5	0	sd
CSMP	2.8	4.1	4.7	5.2	6.0	4.5	2.0
Non-CSMP	2.7	4.0	4.6	5.1	5.9	4.4	2.0

Notes:

1. While the performance for CSMP and Non-CSMP students was nearly identical on items 1-5, the CSMP students did better on items 6 and 7 - the more difficult "multi stage" type - and did less well on the very difficult last item involving fractions.
2. The scale was moderately speeded; 16% of the student did not complete item 7.


Third Grade Class Means, All Sites
x = CSMP Class, . = Non-CSMP Class
SCALE A6: Number Line Labelling
5
4
3
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 4.8
Non-CSMP Classes = 4.5
Probability Level = .06

Analysis of Scale A7

Hints and Problems

	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
Hint: 298 + 168 = 461 298 + 166 = ☐	37	28	.92	458
Hint: 358 + 279 = 637 348 + 279 = ☐	50	48	.82	636 or 537
Hint: 538 + 198 = 736 537 + 198 = ☐	40	37	.90	726 or 636
Hint: 479 + 247 = 726 679 + 247 = ☐	27	23	1.04	746 or 728; 706 or 724
Hint: 165 + 378 = 543 185 + 378 = ☐	31	24	1.05	523; 545 or 563

Correlations

KR 20 Reliability Homogeneity = .77

Kuhlmann Anderson = .51

ETS Cooperative Reading = .59

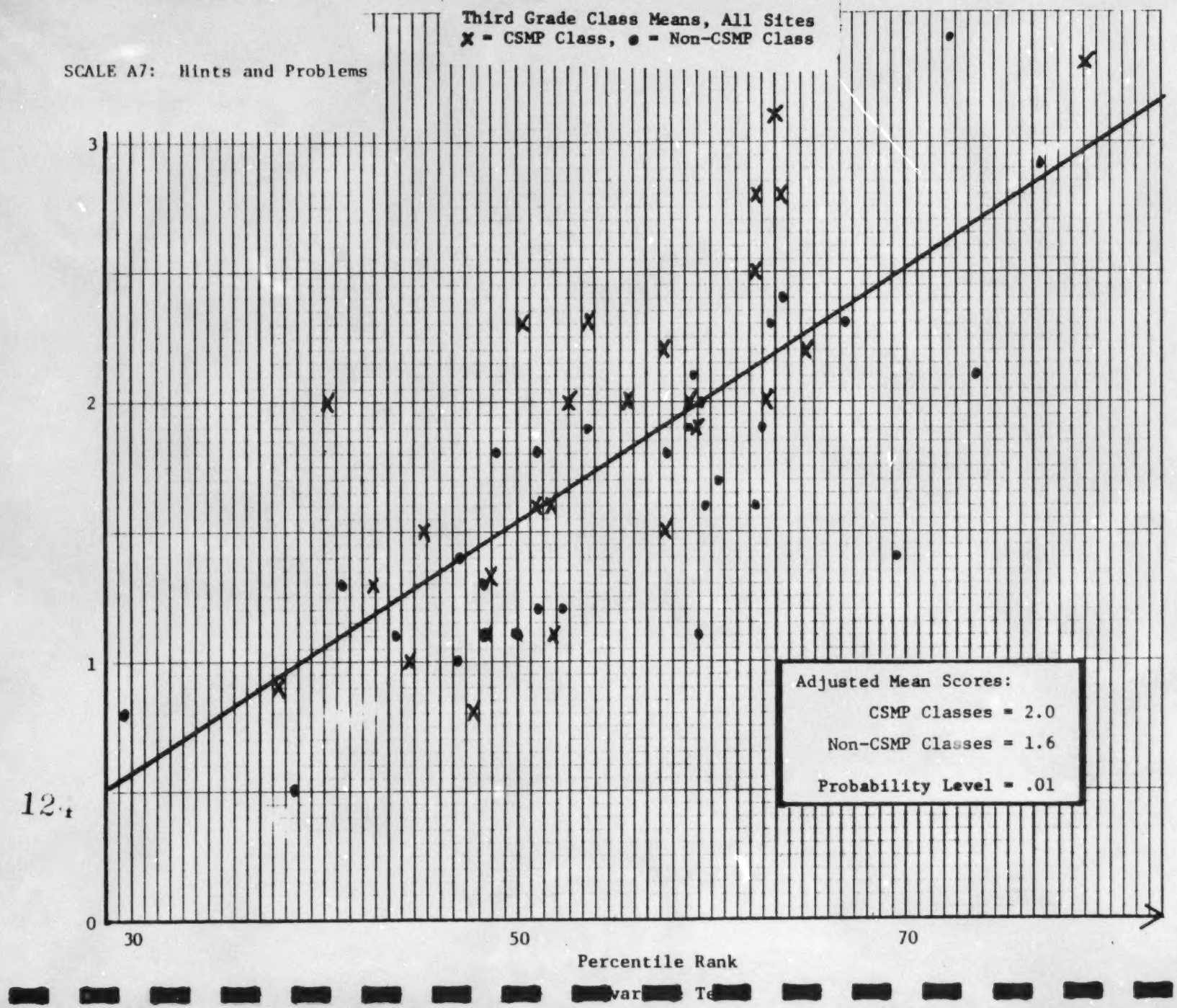
CTBS Computation = .50

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	0.6	1.6	1.9	2.2	3.1	1.9	1.7
Non-CSMP	0.5	1.1	1.5	2.1	2.8	1.6	1.6

Notes:

1. This scale was very hard for many students. About 30% did not get any items correct and another 25% got only one correct. On the other hand about 10% got all five correct.
2. The scale was also moderately speeded; 11% of the students did not complete item 4.
3. About 15% of the wrong answers could be put into one of three categories:
   a) wrong direction (producing an answer for item 1 of 458)
   b) wrong digit change (producing an answer for item 1 of 761 or 491)
   c) both a) and b) (producing an answer for item 1 of 161 or 431)


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE A7: Hints and Problems
3
2
1
0
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 2.0
Non-CSMP Classes = 1.6
Probability Level = .01

Analysis of Scale B1

2 or 5 or 10

Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
General Format				
51 is about ____ times as large as 5				
(Use 2 or 5 or 10)				
1. 51 is about __ times as large as 5	73	66	.50	5 times, 2 times
2. 16 is about __ times as large as 7	66	64	.77	5, 10
3. 65 is about __ times as large as 12	60	56	.72	10(31%), 2
4. 200 is about__ times as large as 21	82	86	.78	5, 2
5. 100 is about__times as large as 19	34	31	.37	10,(58%), 2
6. 98 is about __times as large as 50	57	46	.79	5(36%), 10
7. 45 is about__ times as large as 8	65	61	.70	10, 2
8. 602 is about __ times as large as 298	35	35	.56	5(32%), 10
9. 503 is about __ times as large as 49	63	64	.51	5, 2
10. 59 is about __ times as large as 31	62	61	.73	5, 10

Correlations

KR 20 Reliability Homogeneity = .65

Kuhlmann-Anderson = .50

ETS Cooperative Reading = .51

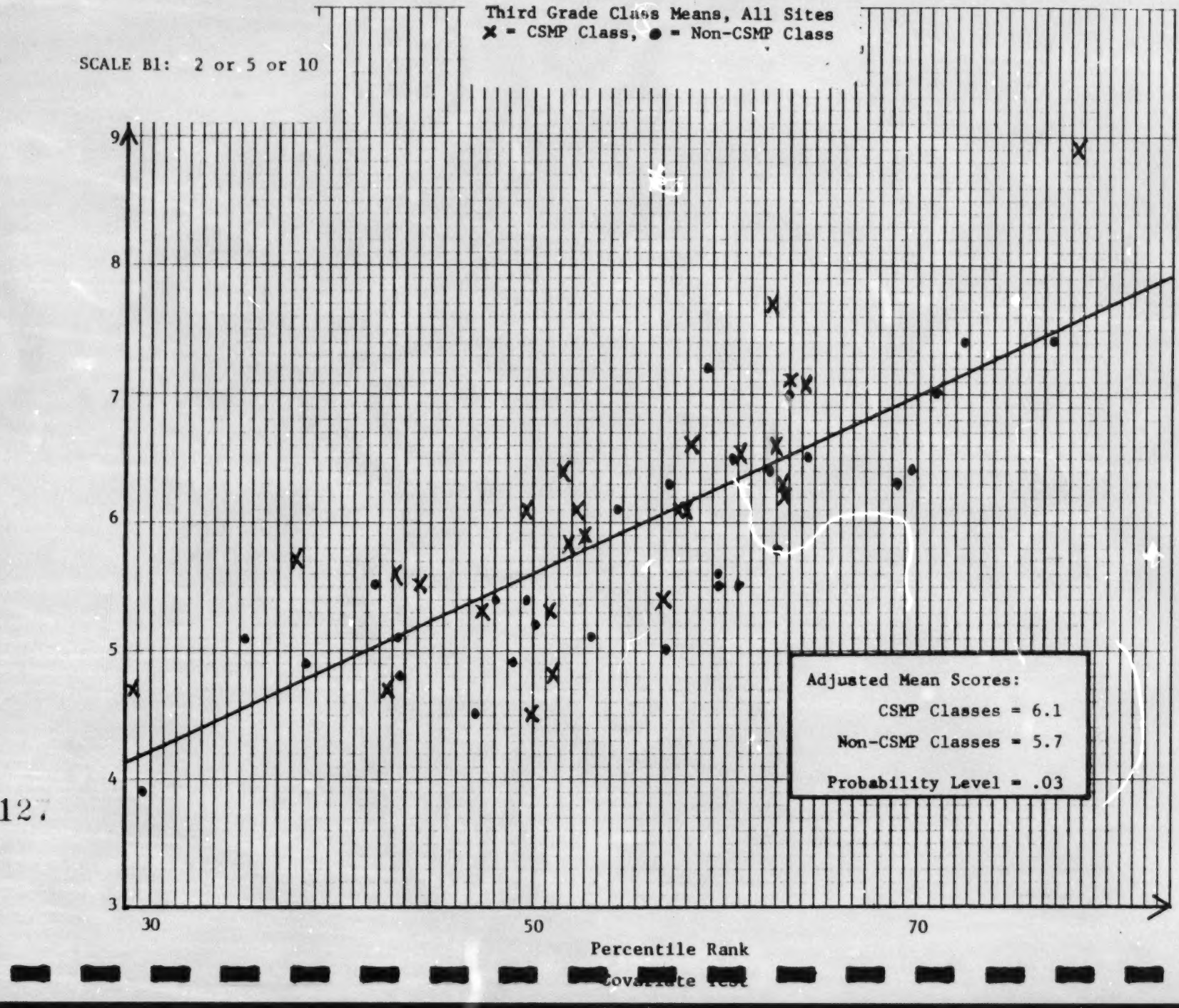
CTBS Computation = .44

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	4.6	5.3	5.8	6.6	7.6	6.0	2.3
Non-CSMP	4.3	5.0	5.7	6.1	7.1	5.7	2.3

Notes:

1. There seemed to be no pattern which described the items on which CSMP students did better.
2. Where "2" was the correct answer, "10" could be considered a "bad" answer and vice versa. Across the seven items where 2 or 10 was correct, the mean percent correct was 62, the mean percent "bad" responses was 11, and the mean percent responding "5" was 27.
3. Across the 3 items where "5" was correct, the mean percent choosing "5" was 51, the mean percent choosing "2" was only 11, and the mean percent choosing "10" was 35! Thus, students were much more likely to err by giving too large a multiplier.



Covariate Test

Analysis of Scale B2
Composite Functions

Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
General Format				
1. 5 ⇒ John ⇒ ☐	89	83	.81	
2. 4 ⇒ Bill ⇒ Mary ⇒ ☐	79	67	.94	1
3. 2 ⇒ John ⇒ John ⇒ ☐	73	61	.92	8,6
4. 0 ⇒ John ⇒ Mary ⇒ Bill ⇒ ☐	68	59	.83	0
5. 12 ⇒ Mary ⇒ Mary ⇒ Mary ⇒ ☐	72	65	.84	9
6. ☐ ⇒ Mary ⇒ 4	68	64	.93	1,2
7. ☐ ⇒ John ⇒ Bill ⇒ 10	49	33	.89	5,2
8. ☐ ⇒ Bill ⇒ Bill ⇒ 8	54	40	.85	4
9. ☐ ⇒ John ⇒ John ⇒ John ⇒ 14	51	38	.80	4,3

Correlations

KR 20 Reliability Homogeneity = .83

Kuhlmann-Anderson = .57

ETS Cooperative Reading = .64

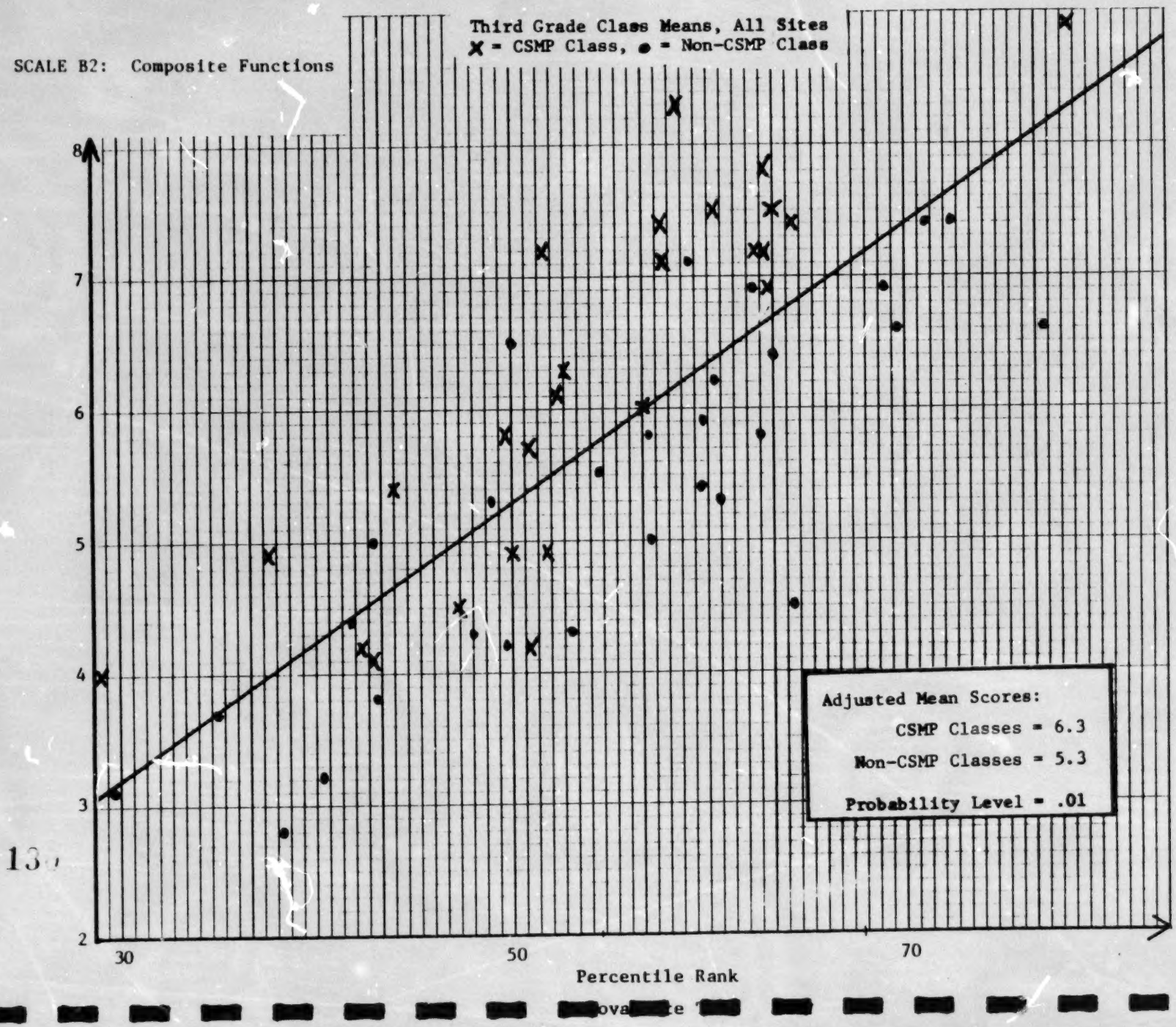
CTBS Computation = .59

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	3.9	5.2	6.4	7.1	8.0	6.0	2.7
Non-CSMP	2.8	4.1	5.5	6.2	7.3	5.2	2.7

Notes:

1. For all items in which 2 names appeared (2 operations to be carried out), the most commo incorrect answer was obtained by ignoring the first name (operation). For example, in #2, 4 ⇒ Mary ⇒ ☐ would produce "1".
2. For CSMP students, 23% got every item correct versus 12% of the Non-CSMP students.


SCALE B2: Composite Functions
Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
8
7
6
5
4
3
2
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 6.3
Non-CSMP Classes = 5.3
Probability Level = .01

Analysis of Scale B3

Equation Fluency

Using only these symbols: + - x ÷ 1 2 3 ( )

	CSMP	Non-CSMP
a) Mean number of correct sentences produced	6.8	6.7
b) Mean number incorrect	0.9	0.9
c) Mean number using "+"	3.3	2.9
d) Mean number using "-"	1.9	1.8
e) Mean number using "x"	3.0	2.9
f) Mean number using more than one operation	1.6	1.2
g) Mean number using parentheses	0.3	0.3

In developing class means, the score on this scale was defined as number of correct sentences produced minus number of incorrect sentences produced (i.e., a) - b) above). Sentences with unallowable symbols (e.g., the "5" in 2 + 3 = 5), were ignored; an average of less than one of these per student was recorded.

Correlations

KR 20 Reliability Homogeneity = NA

Kuhlmann Anderson = .53

ETS Cooperative Reading = .36

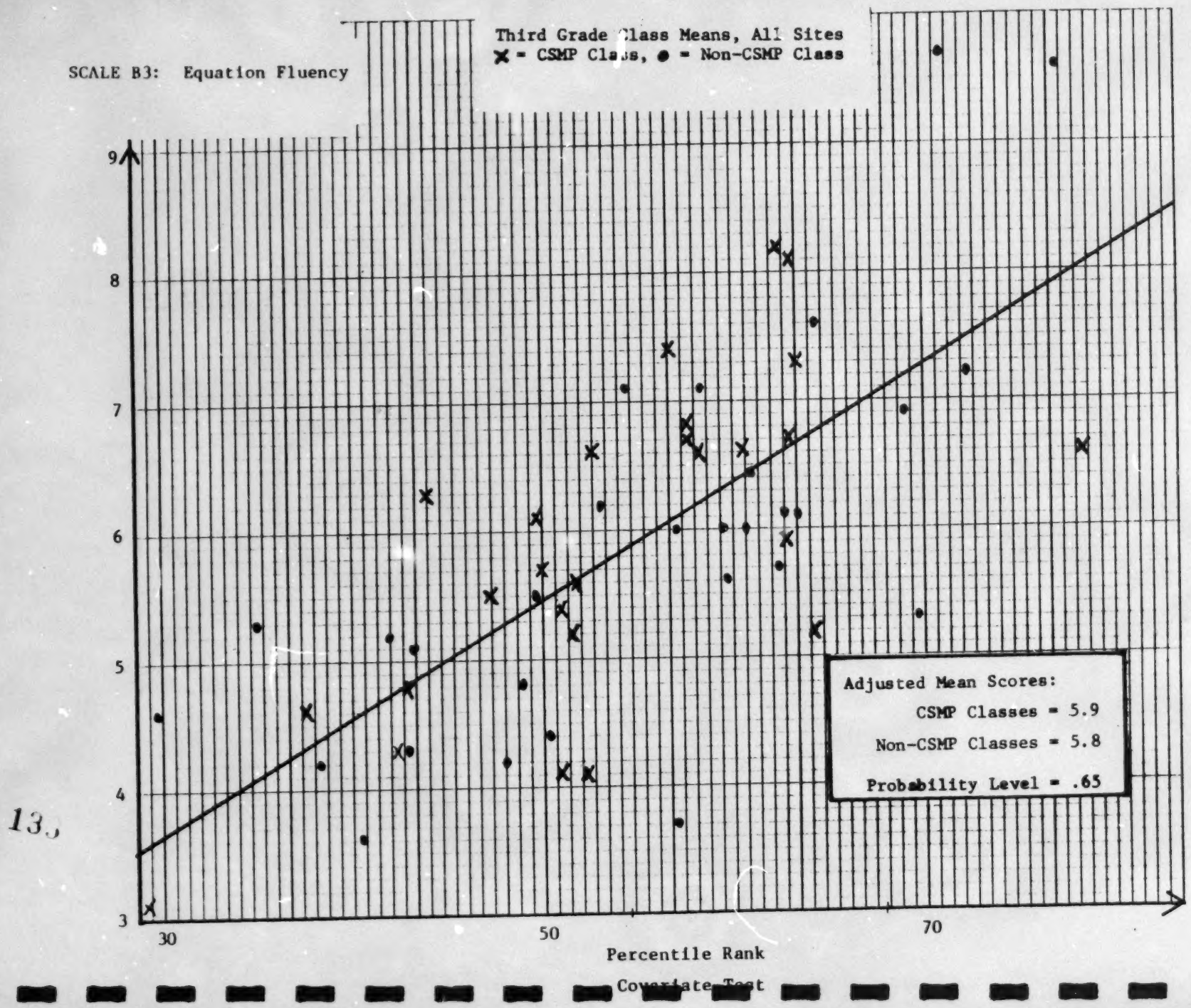
CTBS Computation = .45

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	3.9	4.7	6.3	6.5	7.9	5.8	3.4
Non-CSMP	3.9	4.6	5.9	6.5	7.6	5.8	3.4

Notes:

1. The only places where there was more than 0.1 difference between CSMP and Non-CSMP students were in the number of sentences using "+" and the number using more than operation. CSMP produced more of these two kinds of sentences.


Third Grade Class Means, All Sites
× = CSMP Class, ● = Non-CSMP Class
SCALE B3: Equation Fluency
9
8
7
6
5
4
3
30
50
70
Percentile Rank
Covariate Test
Adjusted Mean Scores:
CSMP Classes = 5.9
Non-CSMP Classes = 5.8
Probability Level = .65

Analysis of Scale B4

Circle the Larger

Items		Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation
General Format 51 + 19	49 + 29			
1. 51 + 19	49 + 29	82	78	.43
2. 700 - 189	710 - 191	58	70	.19
3. $\frac{1}{2}$ of 1000	$\frac{1}{3}$ of 1000	60	34	.62
4. 53 + 19	53 + 23	90	89	.60
5. 7 x 278	7 x 311	82	81	.61
6. 100 - 40	100 - 50	68	48	.68
7. 3 x 162	4 x 160	71	60	.62
8. 460 + 341	490 + 340	74	72	.72
9. $\frac{2}{3}$ of 500	$\frac{1}{6}$ of 500	48	43	.63
10. 500 - 201	500 - 189	61	51	.60
11. 69 + 69 + 69	5 x 69	72	56	.80
12. $\frac{1}{2}$ of 790	2 x 218	40	32	.40
13. 585 + 250	580 + 240	71	72	.53

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	7.2	8.1	8.7	9.5	10.6	8.8	1.5
Non-CSMP	6.4	6.8	7.9	8.5	9.4	7.9	1.6

Correlations

KR 20 Reliability Homogeneity = .62

Kuhlmann-Anderson = .50

ETS Cooperative Reading = .45

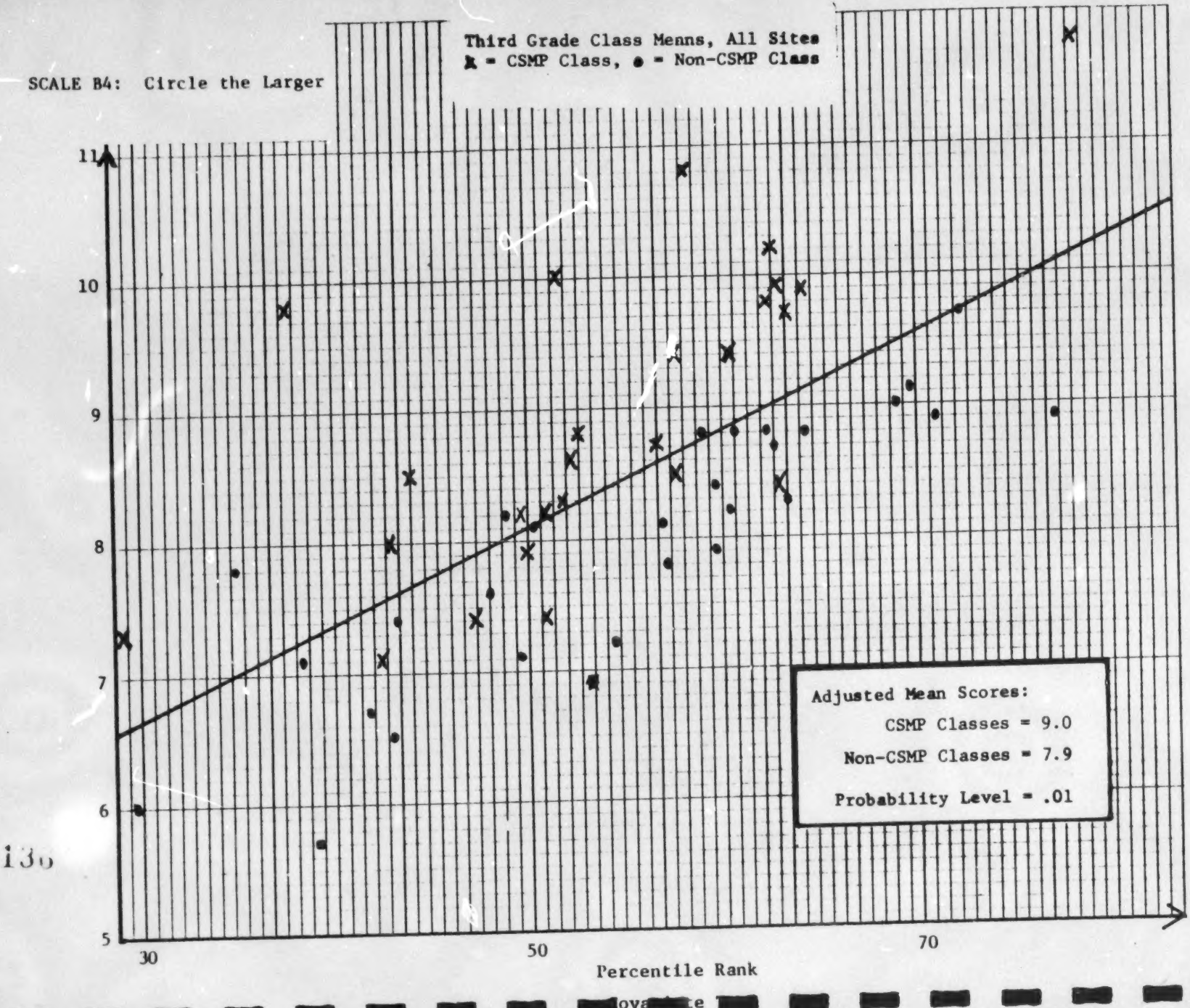
CTBS Computation = .46

Notes:

1. On the 5 addition items there was very little difference between CSMP and Non-CSMP. For the three items on each of subtraction, multiplication, and division, the mean percent correct for CSMP students was about 10% higher than for Non-CSMP students.

2. If one used the simple strategy of choosing whichever side contains the largest numbers (or if equal largest numbers, the second largest number), one would be correct on 5 items and incorrect on 6 (excluding items 3 and 9). For the six items where the strategy doesn't work the mean percents correct were 62 and 49 for CSMP and Non-CSMP students respectively. For items in which the strategy does work the mean percents correct were 75 and 77 respectively. This may help explain the surprising difference on item 2; the very low biserial should also be noted and the data from items 2 and 10 should be compared.

3. The test was slightly speeded, as intended. Eleven percent of the students did not attempt item 12 (and slightly more CSMP than Non-CSMP.


Third Grade Class Means, All Sites
X = CSMP Class, • = Non-CSMP Class
SCALE B4: Circle the Larger
11
10
9
8
7
6
5
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 9.0
Non-CSMP Classes = 7.9
Probability Level = .01

Analysis of Scale B5

Missing Digits

Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
General Format 129 + 2▮▮ = 564 Could 564 be the answer? No, 564 is too small. ☐ Yes, 564 could be right. ☐ No, 564 is too big. ☒				
1. 54 + 3▮▮ = 500	71	70	.53	Could be right, too small
2. 540 + 3▮▮ = 830	26	28	.60	Could be right (41%), too big
3. 364 + 3▮▮ = 700	39	33	.55	Too big (48%), too small
4. 60 - 4▮ = 28	23	22	.58	Could be right (46%), too small
5. 52 - 3▮ = 15	20	16	.45	Too small (67%), too big
6. 550 - 3▮5 = 255	28	23	.54	Could be right (53%), too small
7. 2▮ x 2 = 34	53	43	.61	Too big, too small
8. 13 x ▮ = 239	66	67	.47	Too big, Could be right

Correlations

KR 20 Reliability Homogeneity = .29

Kuhlmann-Anderson = .33

ETS Cooperative Reading = .40

CTBS Computation = .27

Means by Ability Level

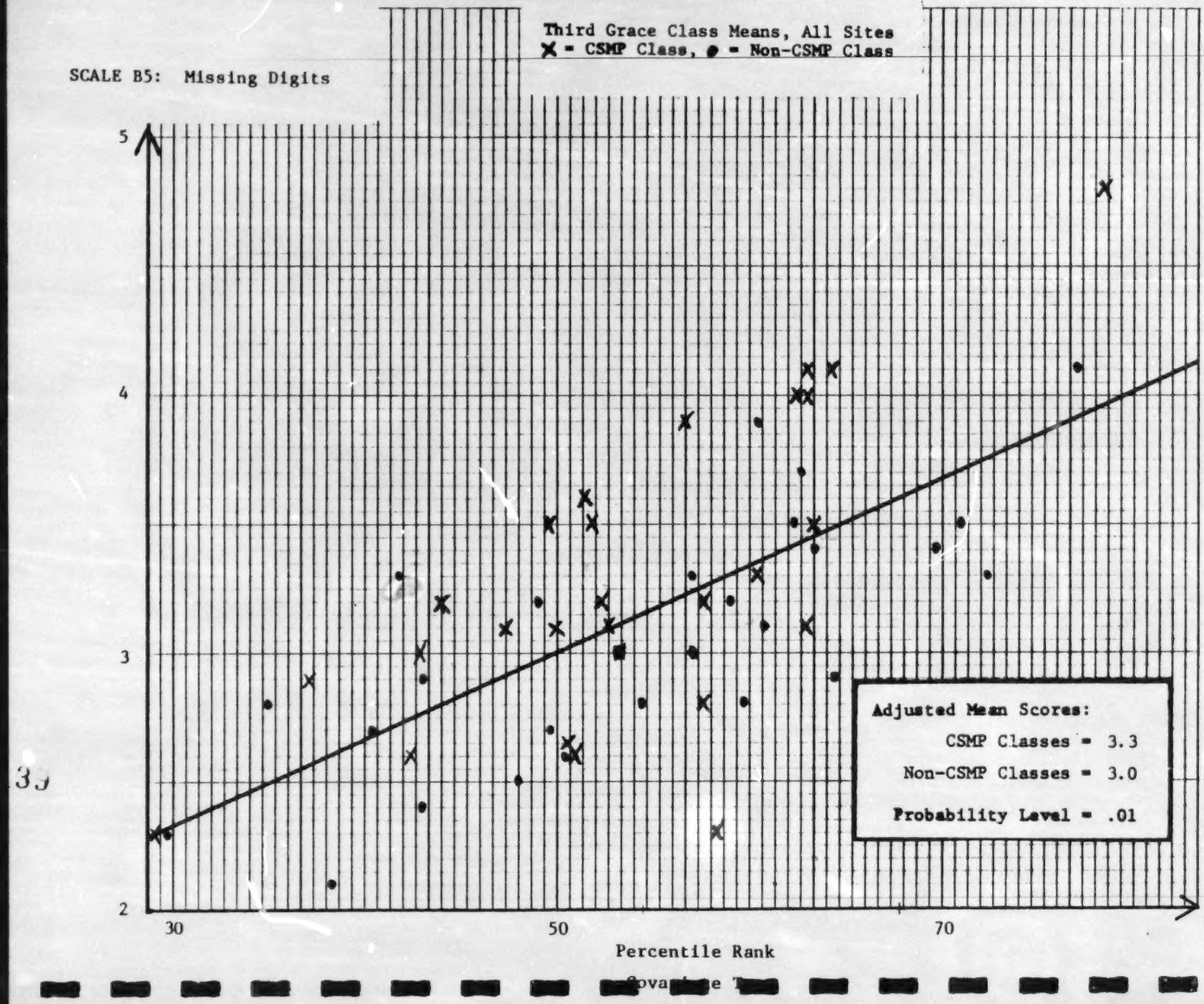
	1	2	3	4	5	0	sd
CSMP	2.7	2.9	3.2	3.6	4.0	3.3	1.5
Non-CSMP	2.5	2.8	2.9	3.0	3.7	3.0	1.4

Notes:

1. The two items where "could be right" was correct (#3, and #5) were harder than the others. The mean percent correct was 30%, 25% for CSMP and Non-CSMP respectively for these two items versus about 48%, 46% for both of the other item types.

2. This scale had the lowest reliability coefficient and, correlations with covariates of all the MANS scales. Note on the graph, next page, that the scores do not cluster particularly well about the regression line.

3. A chance score (blind guessing on each item) would have yielded a score of about 2; about 1/3 of the students scored at this level or below. Analysis of the percentage of students choosing each of the alternatives, does not generally support the argument that much wild guessing occurred.


Third Grace Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE B5: Missing Digits
5
4
3
2
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 3.3
Non-CSMP Classes = 3.0
Probability Level = .01

Analysis of Scale B6

Word Problems "with Rounding"

Items	Percent Correct CSMP	Non-CSMP	Biserial Correlation	Common Errors
. An elevator can't hold more than 5 people. 23 people want to ride to the top floor. How many times will the elevator have to go up? ___	53	50	.85	4, 28, 7
2. Mr. Jones owns a car transporting company. His trucks can carry 3 cars each. He is loading 14 cars onto trucks. How many trucks will he need? ____	54	46	.87	4, 17, 6
3. It takes 4 men to lift a piano. We have 14 men ready to work. How many pianos can they lift at the same time?___	29	30	.76	4, 18, 10, 5
4. There are 18 hamsters in a pet store. The store owner is putting the hamsters into new cages. Each new cage holds no more than 4 hamsters. How many cages will he need? ___	41	41	.82	4, 6
5. A school wants to send as many volleyball teams as possible to a tournament. Each team must have 6 players. There are 21 volleyball players in the school. How many teams can they send? _____	39	42	.84	4

Correlations

KR 20 Reliability Homogeneity = .66

Kuhlmann-Anderson = .43

ETS Cooperative = .53

CTBS Computation = .48

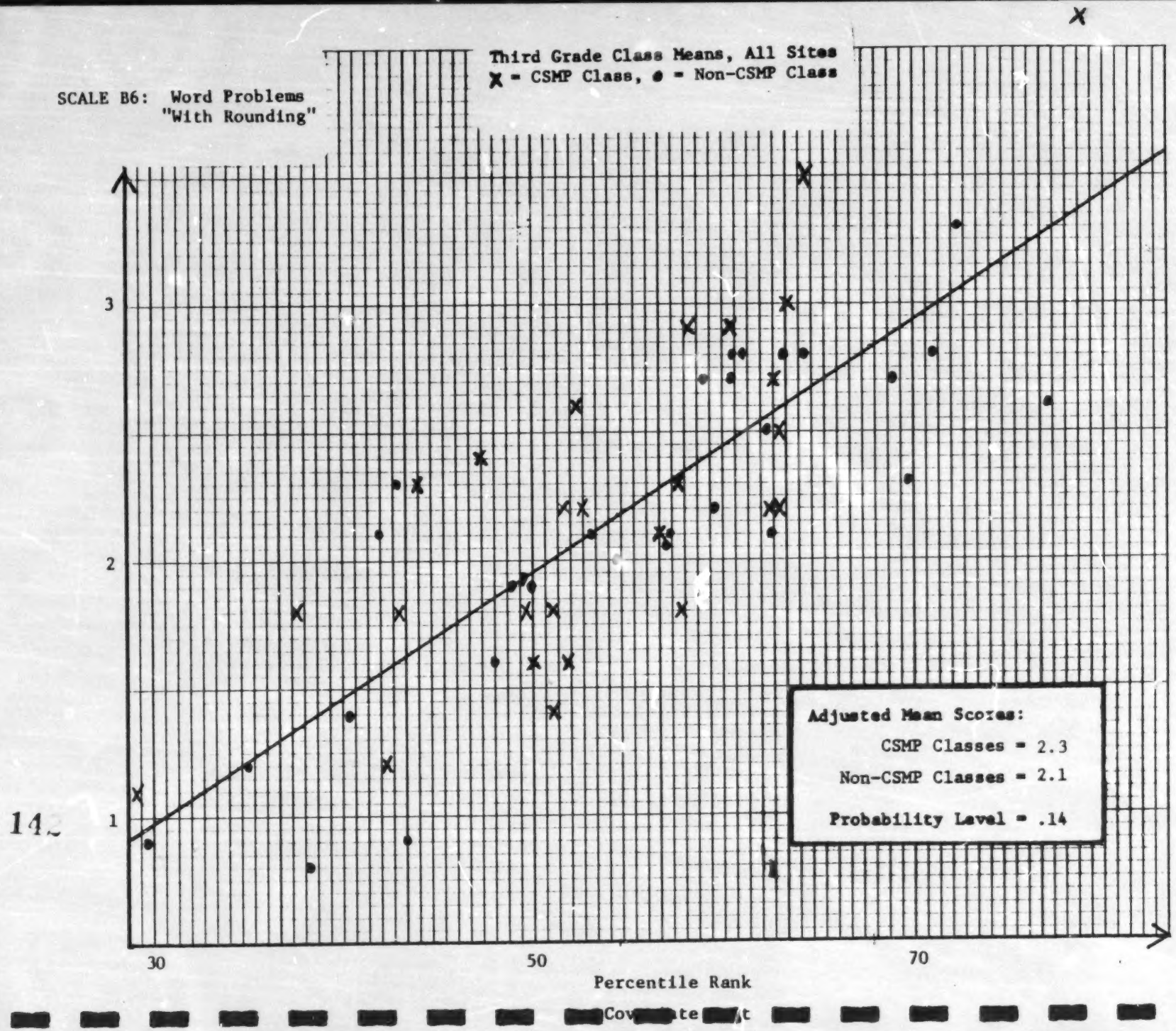
Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	1.4	1.7	2.2	2.5	3.2	2.2	1.6
Non-CSMP	1.1	1.5	2.1	2.4	3.2	2.1	1.6

Notes:

1. The distribution of answers across questions is fairly consistent, with little CSMP-Non-CSMP difference. Across items the mean percent correct was 43 and the most popular wrong answer was always "4" (which was given by an average of 17%). In each item "4" is the wrong answer one gets by ignoring the remainder from the division problem when it should be counted, or vice versa. After "4", about 5% of the responses were the sum of the two numbers, 5% the difference, and 5% one larger and 5% one smaller than the correct answer/best distractor.

2. For both CSMP and Non-CSMP, roughly 1/5 of the students got no items correct and another 1/5 got one item correct.


SCALE B6: Word Problems "With Rounding"
Third Grade Class Means, All Sites
X = CSMP Class, • = Non-CSMP Class
3
2
1
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 2.3
Non-CSMP Classes = 2.1
Probability Level = .14

Analysis of Scale B7

Fractions

Items	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
General Format $\frac{1}{3}$ of 6 = ☐				
1. $\frac{1}{2}$ of 12 = ☐	87	73	.76	
2. $\frac{1}{3}$ of 15 ☐	65	51	.78	
3. $\frac{1}{2}$ of 38 = ☐	33	19	.91	
4. $\frac{1}{2}$ of 230 = ☐	33	18	.98	
5. $\frac{1}{2}$ of ☐ = 8	57	40	.89	4 (36%)
6. $\frac{1}{3}$ of ☐ = 12	24	10	.96	4, 24
7. $\frac{1}{2}$ of ☐ = 28	28	16	.90	14
8. ☐ of 21 = 7	54	41	.73	3

Correlations

KR 20 Reliability Homogeneity = .80

Kuhlmann-Anderson = .58

ETS Cooperative = .65

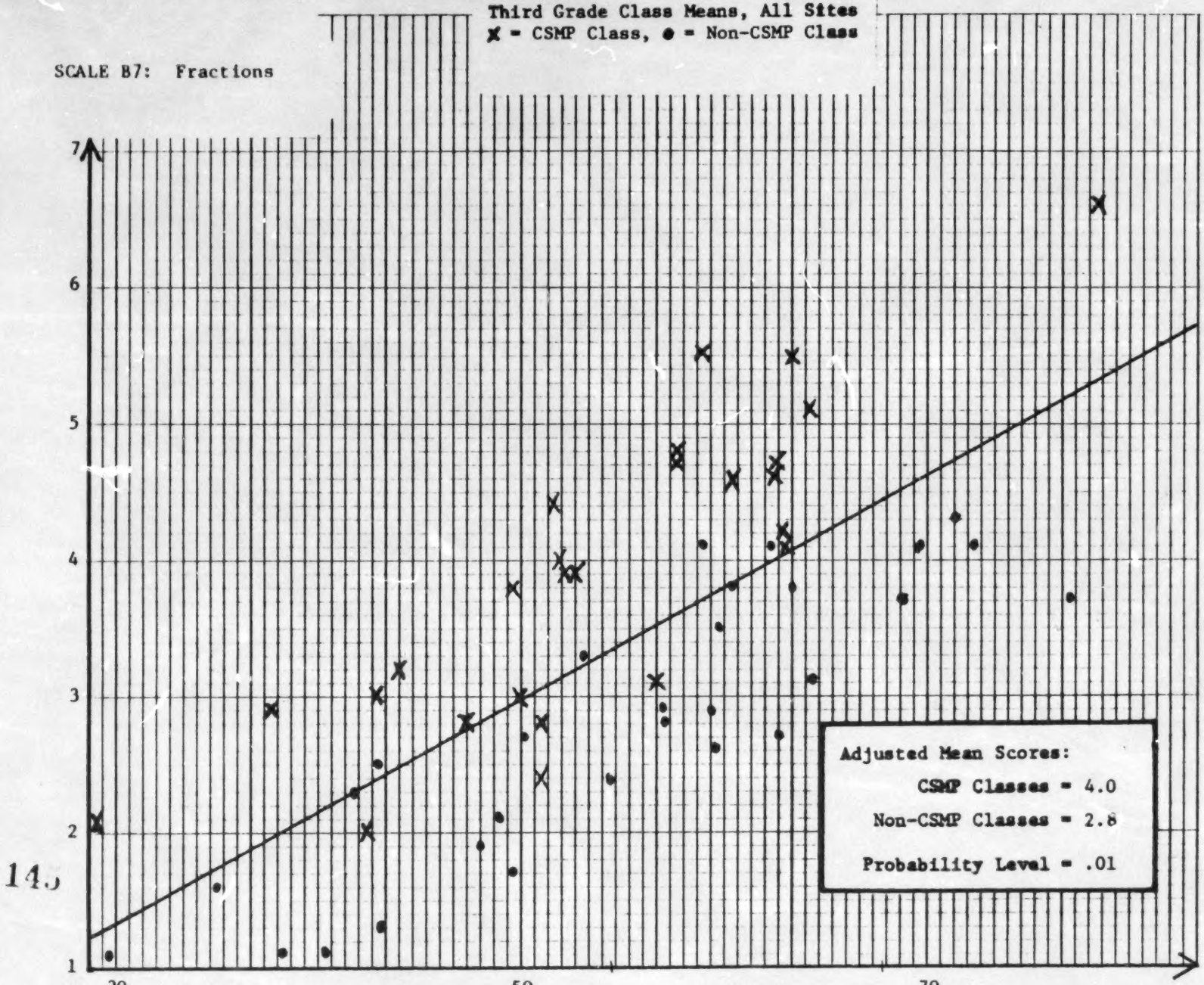
CTBS Computation = .60

Means by Ability Level

	1	2	3	4	5	0	sd
CSMP	1.8	2.8	3.8	4.9	5.9	3.8	2.4
Non-CSMP	1.1	2.0	2.7	3.1	4.4	2.7	2.1

Notes:

1. Of the CSMP students, 21% got at <u>most</u> one item correct out of 8 (versus 35% for Non-CSMP) and 18% got at least 7 items correct (versus 7% for Non-CSMP).


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
SCALE B7: Fractions
7
6
5
4
3
2
1
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 4.0
Non-CSMP Classes = 2.8
Probability Level = .01

Correlations Between MANS A Total and Various Other Tests

MANS B = .85

Comprehensive Test of Basic Skills, Computation = .63

Concepts and Applications = .71

Kuhlmann Anderson Test = .70

ETS Cooperative Primary, Math = .78

ETS Cooperative Primary, Reading = .62

CTBS, Reading Comprehension = .52

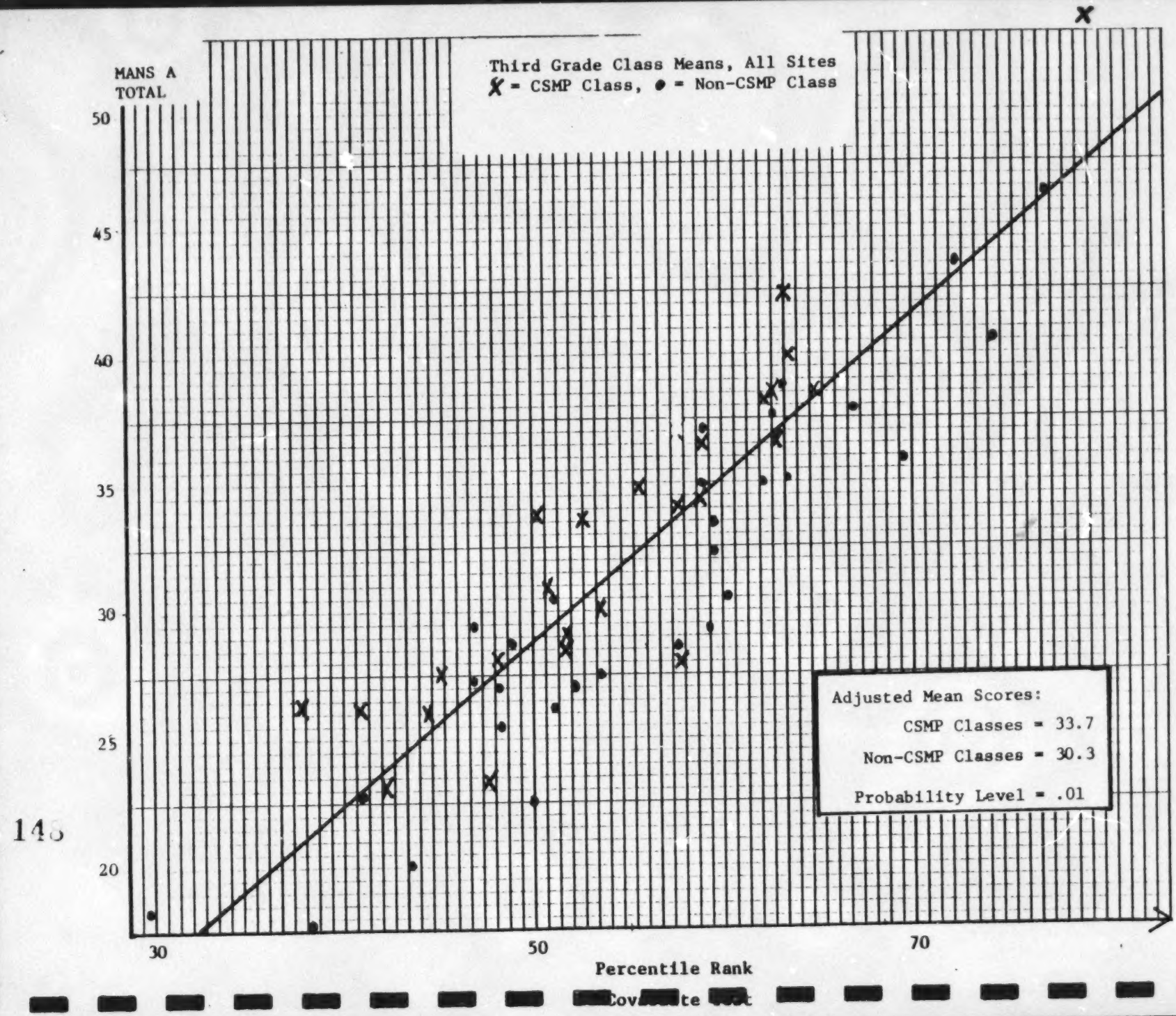
*CTBS, Computation (Spring 76) = .68

*Kuhlmann Anderson (Fall 75) = .71

*1976 version of MANS A = .85

*1976 version of MANS B = .83
(these earlier MANS Tests contained some scales which were very similar to scales in the 1977 MANS)

*Administered while students were in second grade


Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
MANS A
TOTAL
50
45
40
35
30
25
20
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 33.7
Non-CSMP Classes = 30.3
Probability Level = .01

Correlations Between MANS B Total and Various Other Tests

MANS A = .85

Comprehensive Test of Basic Skills, Computation = .67

Concepts and Applications = .75

Kuhlmann Anderson Test = .69

ETS Cooperative Primary, Math = .73

ETS Cooperative Primary, Reading = .59

CTBS, Reading Comprehension = .55

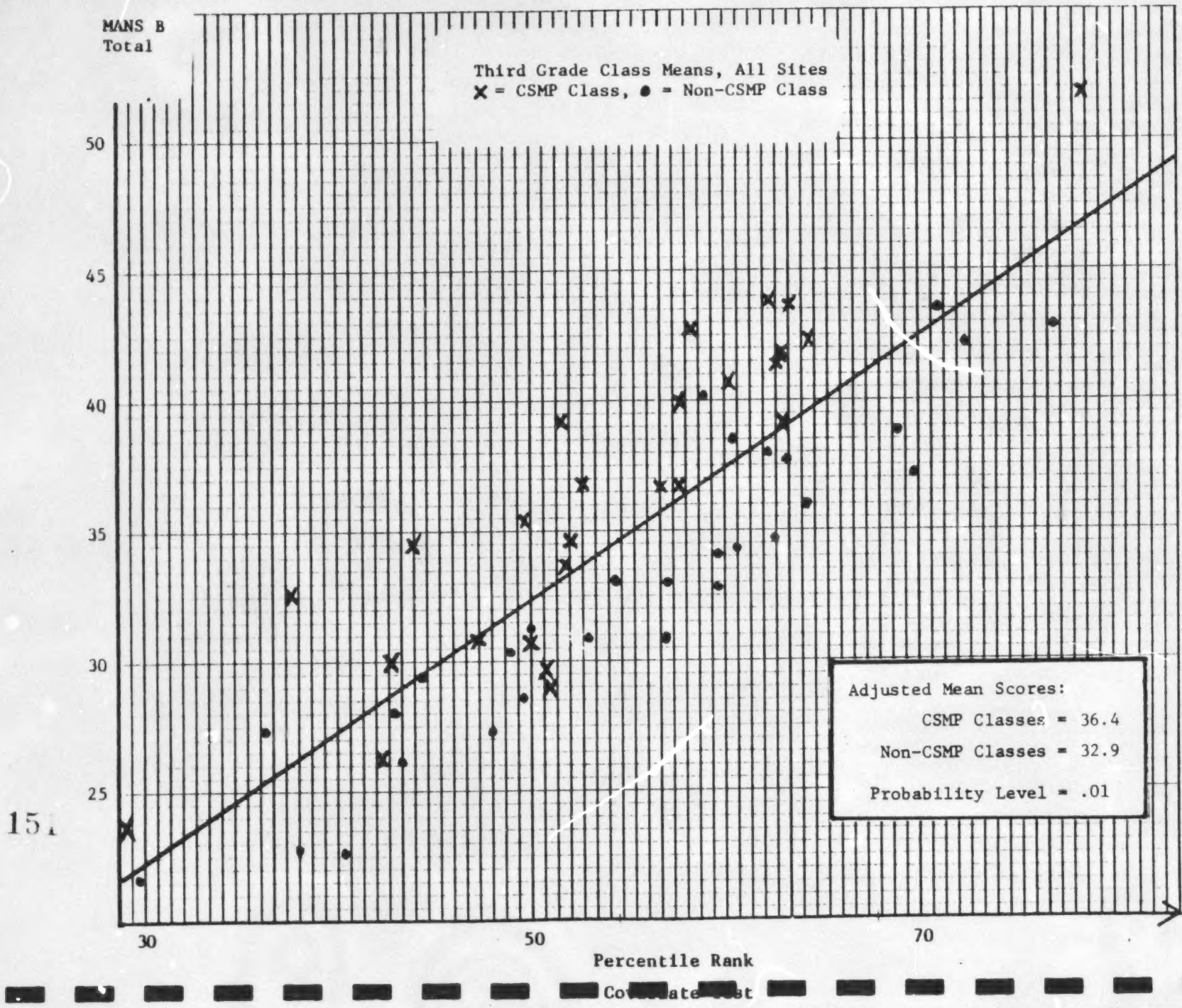
*CTBS, Computation (Spring 76) = .69

*Kuhlmann Anderson (Fall 75) = .72

*1976 version of MANS A = .80

*1976 version of MANS B = .80
(these earlier MANS Tests contained some scales which were very similar to scales in the 1977 MANS)

*Administered while students were in second grade


MANS B
Total
Third Grade Class Means, All Sites
× = CSMP Class, ● = Non-CSMP Class
50
45
40
35
30
25
30
50
70
Percentile Rank
Adjusted Mean Scores:
CSMP Classes = 36.4
Non-CSMP Classes = 32.9
Probability Level = .01